

Superknoten in Sozialen Netzwerken

– eine entropieoptimale Analyse –

Dominic Brenner, Wilhelm Rödder, Friedhelm Kulmann

Diskussionsbeitrag Nr. 485

2013

Diskussionsbeiträge
der Fakultät für Wirtschaftswissenschaft
der FernUniversität in Hagen
Herausgegeben vom Dekan der Fakultät
Alle Rechte liegen bei den Verfassern

Inhaltsverzeichnis

| | | |
|----------|--|-----------|
| 1 | Einleitung | 1 |
| 2 | Wissen über Strukturen Sozialer Netzwerke | 3 |
| 2.1 | Syntax und Semantik | 3 |
| 2.2 | Erwerb und Abruf von Wissen über die Netzstruktur | 4 |
| 3 | Starke Zusammenhangskomponenten (SZK) | 9 |
| 3.1 | Informationstheoretische Charakterisierung von SZK | 9 |
| 3.2 | Kompression von SZK zu Superknoten | 13 |
| 4 | Netzreduktion durch Kompression | 17 |
| 5 | Zusammenfassung und Ausblick | 23 |

Kapitel 1

Einleitung

Die Soziale Netzwerkanalyse (SNA) will Interaktionen zwischen Akteuren in einem Netzwerk beschreiben, die Netzwerkstruktur herausarbeiten, die Vor- oder Nachteilhaftigkeit der Netzwerkzugehörigkeit einzelner Akteure oder von Gruppen solcher Akteure ermitteln und damit schließlich auch Handlungsanweisungen für den Nutzer des Sozialen Netzes (SN) geben. Weitere Beschreibungselemente der SNA findet der Leser bspw. in Scott (1991) oder auch in Rödder u. a. (2012).

Da SN Akteure und Relationen zwischen diesen darstellen, bedient man sich häufig einer graphentheoretischen Repräsentationsform $G = (V, E)$, wobei V die Menge der Akteure und E die Menge der Relationen bedeuten. Graphen oder Hypergraphen sind ein in der Literatur anerkanntes Instrument zur Abstraktion konkreter SN. Die eigentliche SNA wird dann reduziert auf die Analyse dieser Graphen (Barnes, 1969). Naturgemäß ist der Graph nur ein ungenaues Modell der Wirklichkeit; die an diesem Modell gemachten Beobachtungen und die daraus gezogenen Schlüsse sind nur teilweise auf das Original-SN übertragbar.

So werden die Vermaschung des Netzwerkes und die Position eines Akteurs im Netz oft durch graphentheoretische Indizes gemessen: Netzdichte oder Zentralitätsmaße wie Degree-, Closeness-, Betweenness- oder Eigenvector-Centrality. Der Leser wird auf die einschlägige Literatur verwiesen (Newman, 2010; Wasserman u. Faust, 1994). Eine kurze Einführung findet sich auch wiederum in Rödder u. a. (2012).

Dieser Arbeitsbericht setzt auf den Überlegungen in Rödder u. a. (2012) auf und erweitert sie. Erstmals wurde dort die graphentheoretische Analyse durch eine informationstheoretische und hier speziell eine entropiebasierte ersetzt. Für eine bestimmte Form sozialer Kontakte wurden die gerichteten Kanten eines Graphen als Konditionale formuliert und dann die oben genannten Indizes um informationstheoretische Maße erweitert. Dabei kann auch dieses Instrumentarium nur dann Erkenntnisse liefern, wenn es zu dem zu analysierenden SN passt. Borgatti (2005) liefert eine Typologie hinsichtlich der Relationen zwischen Akteuren. Er fokussiert dabei auf solche Netze, in denen (im-)materielle Güter, Neigungen, Verhaltensweisen, Krankheitserreger, Meldungen etc. von Akteur zu Akteur fließen oder vererbt werden. Seine Typologie findet sich in Tabelle 1.1.

In Rödder u. a. (2012) werden die einzelnen Felder erläutert; hier wie dort untersuchen wir Netzstrukturen, die dem Feld [Walks, Parallel duplication] entsprechen. Neigungen oder Attitüden werden von Akteur zu Akteur weitergegeben und verbrauchen sich dabei nicht.

| [Replication] [Trajectory] | Parallel duplication | Serial duplication | Transfer |
|-------------------------------|----------------------|----------------------|------------------|
| Geodesics | <i>No process</i> | Mitotic reproduction | Package delivery |
| Paths | Internet name-server | Viral infection | Mooch |
| Trails | E-mail broadcast | Gossip | Used goods |
| Walks | Attitude influencing | Emotional support | Money exchange |

Tabelle 1.1: Typologie von Relationen nach (Borgatti, 2005)

Will heißen: Verfügt ein Akteur z.B. über eine Information, so gibt er sie voll und ganz an eine direkt von ihm erreichbare Menge von Akteuren weiter; und diese wiederum an andere Mitakteure usw.

Moderne SN umfassen oft in die Million gehende Akteursmengen und noch größere Zahlen an Verbindungen zwischen ihnen. In solchen Fällen ist die Reduktion der Netzstruktur – z.B. durch Zusammenfassung von Akteursgruppen ohne wesentlichen Informationsverlust – ein zentrales Thema. Dieser Frage ist der vorliegende Beitrag gewidmet.

Die Arbeit gliedert sich wie folgt. In Kapitel 2 erfolgt eine kurze Einführung in die entropieoptimale Modellierung von SN. Dazu werden in 2.1 Syntax und Semantik festgelegt und wird in 2.2 der Wissensaufbau beschrieben. Zum besseren Verständnis zeigen wir einige Wissensbasen zu (kleineren) Netzen. Kapitel 3 betrachtet die besondere Rolle von starken Zusammenhangskomponenten (SZK) im entropiebasierten Netz. In 3.1 wird ein Netz mittlerer Größe vorgestellt, in dem mehrere SZK vorliegen, und es werden ihre informationstheoretischen Charakteristika erarbeitet. In 3.2 berichten wir über die Kompression von SZK zu Superknoten und die informationstheoretischen Konsequenzen der Kompression. Kapitel 4 schließlich zeigt das Reduktionspotenzial durch Kompression in großen Netzen auf. Kapitel 5 fasst die Ergebnisse zusammen und wagt einen Blick auf zukünftige Forschungsarbeiten.

Kapitel 2

Wissen über Strukturen Sozialer Netzwerke

2.1 Syntax und Semantik

In Rödder u. a. (2012) wird ausführlich über den Aufbau von Wissen über konditionale Strukturen im Allgemeinen und im Besonderen für SN berichtet. Im Kontext des vorliegenden Berichts wählen wir eine leicht modifizierte und reduzierte Form der Darstellung – eben auf die jetztige Thematik zugeschnitten. Zentrale Bausteine der Wissensrepräsentation sind

- eine Variablenmenge $V = \{V_1, \dots, V_L\}$ mit Werten oder Ausprägungen $V_l = v_l$ mit $v_l = 1$ oder 0 für $l \in \{1, \dots, L\}$,
- Konditionale $V_l = 1 \mid V_k = 1$ für einige $k \neq l \in \{1, \dots, L\}$.

Die Semantik einer Variablen V_l ist die einer Proposition: Akteur l verfügt über ein (im-)materielles Gut, hat eine Neigung oder eine Verhaltensweise (1) oder nicht (0). Die Semantik eines Konditionals ist die der potenziellen Weitergabe eines solchen Artefakts von Akteur k zu l . Das Konditional $V_l = 1 \mid V_k = 1$ erlaubt und erzwingt die Weitergabe des Gutes von k zu l , falls k darüber verfügt. Durch die Angabe der Konditionale wird also gleichsam die potenzielle Flussstruktur des Gutes festgelegt. Tabelle 2.1 zeigt die Menge der 2^L möglichen Zustände des Netzes, ergänzt um 2^L Wahrscheinlichkeiten.

Der Schlüssel zum Verständnis der Modellierung der Netzwerkstruktur sind die Wahrscheinlichkeiten! Ihre genaue Festlegung wird im Folgeabschnitt erläutert, eine notwendige Bedingung für ihre Festlegung soll jedoch hier schon erwähnt werden.

Tabelle 2.1: Zustandsraum und Wahrscheinlichkeiten eines SN

| $V_1 V_2 \dots V_L$ | Wkt. |
|---------------------|----------------|
| 1 1 ... 1 | $p_{1\dots 1}$ |
| 1 1 ... 0 | $p_{1\dots 0}$ |
| \vdots | \vdots |
| 0 0 ... 0 | $p_{0\dots 0}$ |

$$\frac{V_1 \dots V_l \dots V_k \dots V_L}{\dots 0 \dots 1 \dots} \Bigg| \begin{array}{l} \text{Wkt.} \\ p \dots 0 \dots 1 \dots \end{array} \quad (2.1)$$

Betrachten Sie den Zustand in (2.1) und vollziehen Sie nach: Soll das Netz die potenzielle Weitergabe des Gutes von k zu l modellieren, ist also $V_l = 1 \mid V_k = 1$ gefordert, ist der Zustand in (2.1) unmöglich. Denn er würde ja bedeuten: V_k hat das Gut, V_l aber nicht. Diese Unmöglichkeit wird durch die Wahrscheinlichkeit $p \dots 0 \dots 1 \dots = 0.0$ ausgedrückt. Es zeichnet sich bereits jetzt ab, dass auf der Menge der 2^L Zustände eine Wahrscheinlichkeitsverteilung festzulegen ist, die zumindest die unmöglichen Zustände mit 0.0 belegt. Alles Weitere im Folgeabschnitt. Wenn unmissverständlich, bedienen wir uns zur Visualisierung von Konditionalen auch gerichteter Kanten oder Pfeile. $V_k \rightarrow V_l$ oder gelegentlich auch einfach $k \rightarrow l$ steht dann für das Konditional $V_l = 1 \mid V_k = 1$. Trotz dieser anschaulichen Darstellung muss stets beachtet werden, dass im Wissensverarbeitungsprozess der vorliegenden Arbeit der Pfeil die Semantik eines Konditionals hat.

2.2 Erwerb und Abruf von Wissen über die Netzstruktur

Bekannt seien nunmehr alle erwünschten potenziellen Güterflüsse im Netz, ausgedrückt als eine Menge von I sicheren Konditionalen der Form

$$\mathcal{R} = \{V_{l_i} = 1 \mid V_{k_i} = 1[1.0] \text{ für einige } l_i \neq k_i \text{ und } i = 1, \dots, I\}. \quad (2.2)$$

Jedes Konditional bedeutet hierbei: Falls Akteur k_i über das Gut verfügt, so auch – mit Wahrscheinlichkeit 1.0 – Akteur l_i . Zur Erzeugung einer Verteilung auf dem Zustandsraum löst man nun die Aufgabe

$$P^* = \arg \min R(V; Q, P^0) \text{ s.d. } Q \models \mathcal{R}. \quad (2.3)$$

P^0 ist die Gleichverteilung auf dem Zustandsraum, Q die variable Verteilung in (2.3), sie ist die Entscheidungsvariable. $\arg \min$ bestimmt als P^* die Verteilung, die (2.3) löst, R ist die Relative Entropie von P^0 zu Q auf V . Die Aufgabe (2.3) errechnet die Verteilung $Q = P^*$ so, dass alle sicheren Konditionale in \mathcal{R} erfüllt sind – $Q \models \mathcal{R}$ – und dabei minimalen informationstheoretischen Abstand von P^0 hat. Hierbei weist P^0 allen Zuständen des Netzes die gleiche Wahrscheinlichkeit $\frac{1}{2^L}$ zu, sie „kennt“ die geforderte Netzstruktur (noch) nicht. Durch Lösen von (2.3) wird sie zu P^* abgeändert und über die Flussstruktur „informiert“. Bekanntlich kann die Zielfunktion in (2.3) durch $P^* = \arg \max H(V; Q)$ ersetzt werden, wobei H die Entropie der Verteilung Q auf V ist. Aus diesem Grund wird die Wissensakquisition gemäß (2.3) auch oft entropiebasiert genannt. Maximal ist die Entropie der Verteilung auf dem Zustandsraum dann, wenn alle Wahrscheinlichkeiten entweder 0.0 – siehe oben – oder gleich sind. Jedes Lehrbuch über Informationstheorie hält den Nachweis dieser Aussage bereit, das folgende Beispiel verdeutlicht die Zusammenhänge.

Beispiel 1. Zu den Netzen i) $V_1 \rightarrow V_2$, ii) $V_1 \rightarrow V_2 \rightarrow V_3$, iii) $V_2 \leftarrow V_1 \rightarrow V_3$ und iv) $V_1 \rightarrow V_3 \leftarrow V_2$ werden nun die Zustände und die Lösungen von (2.3) angegeben.

| Zu i) | V_1 | V_2 | P^* |
|-------|-------|-------|---------------|
| | 1 | 1 | $\frac{1}{3}$ |
| | 1 | 0 | 0.0 |
| | 0 | 1 | $\frac{1}{3}$ |
| | 0 | 0 | $\frac{1}{3}$ |

| Zu ii) | V_1 | V_2 | V_3 | P^* | Zu iii) | V_1 | V_2 | V_3 | P^* | Zu iv) | V_1 | V_2 | V_3 | P^* |
|--------|-------|-------|-------|---------------|---------|-------|-------|-------|---------------|--------|-------|-------|-------|---------------|
| | 1 | 1 | 1 | $\frac{1}{4}$ | | 1 | 1 | 1 | $\frac{1}{5}$ | | 1 | 1 | 1 | $\frac{1}{5}$ |
| | 1 | 1 | 0 | 0.0 | | 1 | 1 | 0 | 0.0 | | 1 | 1 | 0 | 0.0 |
| | 1 | 0 | 1 | 0.0 | | 1 | 0 | 1 | 0.0 | | 1 | 0 | 1 | $\frac{1}{5}$ |
| | 1 | 0 | 0 | 0.0 | | 1 | 0 | 0 | 0.0 | | 1 | 0 | 0 | 0.0 |
| | 0 | 1 | 1 | $\frac{1}{4}$ | | 0 | 1 | 1 | $\frac{1}{5}$ | | 0 | 1 | 1 | $\frac{1}{5}$ |
| | 0 | 1 | 0 | 0.0 | | 0 | 1 | 0 | $\frac{1}{5}$ | | 0 | 1 | 0 | 0.0 |
| | 0 | 0 | 1 | $\frac{1}{4}$ | | 0 | 0 | 1 | $\frac{1}{5}$ | | 0 | 0 | 1 | $\frac{1}{5}$ |
| | 0 | 0 | 0 | $\frac{1}{4}$ | | 0 | 0 | 0 | $\frac{1}{5}$ | | 0 | 0 | 0 | $\frac{1}{5}$ |

Bitte vollziehen Sie nach, dass die Wahrscheinlichkeiten in P^* stets 0.0 oder gleich sind und dass ferner die den Pfeilen entsprechenden Konditionale nie verletzt sind: Ist ein $V_k = 1$, so ist $V_l = 0$ unmöglich, falls es einen Pfad von k nach l gibt.

Soweit der Wissenserwerb über mögliche Flüsse von Artefakten im Netz. Anzumerken ist noch, dass die Konditionalstruktur des Netzes transitiv ist. Falls $V_l | V_k [1.0]$ und $V_j | V_l [1.0]$ gelten, so gilt in P^* : $P^*(V_j | V_k) = 1.0$. Der Nachweis ist schnell erbracht und wird hier ausgespart.

Das aufgebaute Wissensmodell ist rein konditionaler Natur: wenn, dann. Noch ist nicht bekannt, ob und welcher Akteur über das (im-)materielle Gut verfügt bzw. die Neigung oder Verhaltensweise zeigt, die nun weitergegeben werden kann und muss. Wie viele und welche Mitakteure jemand erreicht, bestimmt sein aktives Potenzial im Netz, von wie vielen und welchen er erreicht wird, bestimmt sein passives Potenzial. Solchen Fragen ist der Rest dieses Abschnitts gewidmet.

Wird Akteur k in den Zustand versetzt, über das Artefakt zu verfügen oder nicht (!), nennen wir das Evidenzieren: Akteur k 's Zustand wird evident. Die mathematische Entsprechung dieses Vorgangs ist die Aufgabe (2.4):

$$P^{**}(\overline{P^{**}}) = \arg \min R(V; Q, P^*) \text{ s.d. } Q \models V_k = 1[1.0] \text{ (} Q \models V_k = 0[1.0]\text{)}. \quad (2.4)$$

So werden bei Evidenzierung von $V_k = 1$ alle von ihm erreichbaren Akteure ebenfalls in den Genuss des Artefakts kommen und in den Zustand 1 überführt. So können bei Evidenzierung von $V_k = 0$ alle ihn erreichenden Akteure ebenfalls nicht im Genuss des Artefakts sein und werden in den Zustand 0 überführt.

Beispiel 2. (Fortsetzung von Beispiel 1)

Es werden ausgewählte Evidenzierungen zu den Fällen ii), iii) und iv) vorgenommen und die jeweiligen Verteilungen P^{**} sowie die Randverteilungen auf den Variablen angegeben; zur Berechnung vgl. Rödder u. Kern-Isberner (1997); Meyer (1998, S. 64).

| Zu ii) a. Evidenzierung $V_1 = 1$ | V_1 | V_2 | V_3 | P^{**} |
|-----------------------------------|-------|-------|-------|----------|
| | 1 | 1 | 1 | 1.0 |
| | 1 | 1 | 0 | 0.0 |
| | 1 | 0 | 1 | 0.0 |
| | 1 | 0 | 0 | 0.0 |
| | 0 | 1 | 1 | 0.0 |
| | 0 | 1 | 0 | 0.0 |
| | 0 | 0 | 1 | 0.0 |
| | 0 | 0 | 0 | 0.0 |

$$P^{**}(V_1 = 1) = 1.0, P^{**}(V_2 = 1) = 1.0, P^{**}(V_3 = 1) = 1.0$$

b. Evidenziierung $V_2 = 1$

| V_1 | V_2 | V_3 | P^{**} |
|-------|-------|-------|---------------|
| 1 | 1 | 1 | $\frac{1}{2}$ |
| 1 | 1 | 0 | 0.0 |
| 1 | 0 | 1 | 0.0 |
| 1 | 0 | 0 | 0.0 |
| 0 | 1 | 1 | $\frac{1}{2}$ |
| 0 | 1 | 0 | 0.0 |
| 0 | 0 | 1 | 0.0 |
| 0 | 0 | 0 | 0.0 |

$$P^{**}(V_2 = 1) = 1.0, P^{**}(V_3 = 1) = 1.0$$

c. Evidenziierung $V_3 = 1$

| V_1 | V_2 | V_3 | P^{**} |
|-------|-------|-------|---------------|
| 1 | 1 | 1 | $\frac{1}{3}$ |
| 1 | 1 | 0 | 0.0 |
| 1 | 0 | 1 | 0.0 |
| 1 | 0 | 0 | 0.0 |
| 0 | 1 | 1 | $\frac{1}{3}$ |
| 0 | 1 | 0 | 0.0 |
| 0 | 0 | 1 | $\frac{1}{3}$ |
| 0 | 0 | 0 | 0.0 |

$$P^{**}(V_3 = 1) = 1.0$$

Zu iii) a. Evidenziierung $V_1 = 1$

| V_1 | V_2 | V_3 | P^{**} |
|-------|-------|-------|----------|
| 1 | 1 | 1 | 1.0 |
| 1 | 1 | 0 | 0.0 |
| 1 | 0 | 1 | 0.0 |
| 1 | 0 | 0 | 0.0 |
| 0 | 1 | 1 | 0.0 |
| 0 | 1 | 0 | 0.0 |
| 0 | 0 | 1 | 0.0 |
| 0 | 0 | 0 | 0.0 |

$$P^{**}(V_1 = 1) = 1.0, P^{**}(V_2 = 1) = 1.0, P^{**}(V_3 = 1) = 1.0$$

b. Evidenziierung $V_2 = 1$

| V_1 | V_2 | V_3 | P^{**} |
|-------|-------|-------|---------------|
| 1 | 1 | 1 | $\frac{1}{3}$ |
| 1 | 1 | 0 | 0.0 |
| 1 | 0 | 1 | 0.0 |
| 1 | 0 | 0 | 0.0 |
| 0 | 1 | 1 | $\frac{1}{3}$ |
| 0 | 1 | 0 | $\frac{1}{3}$ |
| 0 | 0 | 1 | 0.0 |
| 0 | 0 | 0 | 0.0 |

$$P^{**}(V_2 = 1) = 1.0$$

Zu iv) a. Evidenzierung $V_1 = 1$

| V_1 | V_2 | V_3 | P^{**} |
|-------|-------|-------|---------------|
| 1 | 1 | 1 | $\frac{1}{2}$ |
| 1 | 1 | 0 | 0.0 |
| 1 | 0 | 1 | $\frac{1}{2}$ |
| 1 | 0 | 0 | 0.0 |
| 0 | 1 | 1 | 0.0 |
| 0 | 1 | 0 | 0.0 |
| 0 | 0 | 1 | 0.0 |
| 0 | 0 | 0 | 0.0 |

$$P^{**}(V_1 = 1) = 1.0, P^{**}(V_3 = 1) = 1.0$$

b. Evidenzierung $V_3 = 1$

| V_1 | V_2 | V_3 | P^{**} |
|-------|-------|-------|---------------|
| 1 | 1 | 1 | $\frac{1}{4}$ |
| 1 | 1 | 0 | 0.0 |
| 1 | 0 | 1 | $\frac{1}{4}$ |
| 1 | 0 | 0 | 0.0 |
| 0 | 1 | 1 | $\frac{1}{4}$ |
| 0 | 1 | 0 | 0.0 |
| 0 | 0 | 1 | $\frac{1}{4}$ |
| 0 | 0 | 0 | 0.0 |

$$P^{**}(V_3 = 1) = 1$$

Die Randverteilungen auf den Variablen zeigen das gewünschte Ergebnis. Die erreichbaren Variablen nehmen bei der jeweiligen Evidenzierung den Wert 1 an. Die Berechnung der Verteilungen $\overline{P^{**}}$ und die entsprechenden Ränder auf den Variablen zeigen ebenfalls die gewünschten Ergebnisse. Sie werden hier ausgespart.

Die berechneten Verteilungen P^{**} des Beispiels zeigen also den gewünschten Effekt: Sie übertragen die Evidenzierung $V_k = 1[1.0]$ auf alle von k erreichbaren Akteure und die Evidenzierung $V_k = 0[1.0]$ auf alle den k erreichenden Akteure. Wäre das das Ziel unserer Überlegungen, wäre der Wissensverarbeitungsprozess (2.3) und (2.4) natürlich zu aufwändig; einfache Markierungsalgorithmen würden das Gleiche leisten. Die Vorteilhaftigkeit des Prozesses liegt in den aus den Verteilungen P^* und P^{**} ableitbaren informationstheoretischen Maßen, deren Berechnung benutzerfreundlich in der Expertensystemshell SPIRIT (Rödder, 2009) möglich ist. Die in Rödder u. a. (2012) abgeleiteten Maße und ihre Bedeutung für die Netzanalyse werden hier kurz wiederholt, für eine genauere Darstellung und für mathematische Beweise vergleiche man die Originalliteratur.

- Wird $V_k = 1[1.0]$ wie in (2.4) evidenziert, so misst $R(V; P^{**}, P^*)$ die Information in [bit], die dem Netz durch die Evidenzierung zugeflossen ist. $R(V; P^{**}, P^*)$ ist also gleichsam der globale(!) Wissenszuwachs – von Akteur k ausgelöst; R misst das Diffusionspotenzial (des Akteurs).
- $R(V; P^{**}, P^*)$ ist gleich $-\log_2 P^*(V_k = 1)$. (2.5)

Bereits vor(!) Evidenzierung ist k 's Diffusionspotenzial ablesbar! P^* „kennt“ die Diffusionspotenziale aller Akteure – und diese Potenziale sind in SPIRIT benutzerfreundlich dargestellt. Es bedarf also keines Abzählens von Pfeilen auf Erreichbarkeitspfaden; die Shell hält das Potenzial bereit. Anzumerken ist noch, dass der Wert

$-\log_2 \mathbb{P}^*(V_k = 1)$ nicht nur die Erreichbarkeit „zählt“, sondern auch die Positionen der erreichbaren Akteure im Netz berücksichtigt; mehr dazu wiederum in Rödter u. a. (2012).

- Wird $V_k = 0[1.0]$ wie in (2.4) evidenziert, so misst $R(V; \overline{\mathbb{P}^{**}}, \mathbb{P}^*)$ die Information in [bit], die dem Netz durch die Evidenzierung zugeflossen ist. Es misst die Lage der Akteure, von denen k erreichbar ist. Die oben genannten Autoren nennen es das Rezeptionspotenzial von k .
- $R(V; \overline{\mathbb{P}^{**}}, \mathbb{P}^*)$ ist gleich $-\log_2 \mathbb{P}^*(V_k = 0)$. (2.6)

Auch hier sind die Rezeptionspotenziale aller Akteure bereits in \mathbb{P}^* bekannt und werden in SPIRIT benutzerfreundlich zur Verfügung gestellt.

Bisher wurden kleinere Beispiele von *zyklenfreien* Netzen betrachtet; in vielen Anwendungen enthalten die Netze jedoch starke Zusammenhangskomponenten (SZK), in denen jeder Akteur von jedem über einen Pfad gerichteter Kanten – Konditionale – erreichbar ist. Die informationstheoretische Charakterisierung solcher Netze ist Gegenstand der beiden folgenden Kapitel.

Kapitel 3

Starke Zusammenhangskomponenten (SZK) unter maximaler Entropie

3.1 Informationstheoretische Charakterisierung von SZK

Zur Einführung dieses Abschnitts betrachten wir das modifizierte Newcomb-Netz wie in Abbildung 3.1. Für eine genauere Beschreibung des Netzes siehe Newcomb (1961); Trappmann u. a. (2005) oder wiederum Rödder u. a. (2012). Die modifizierte Semantik des Netzes jedoch möge die einer potenziellen Weitergabe von Neigungen, Informationen oder Botschaften sein; Akteur 18 ist isoliert. Die Konditionalstruktur gemäß Aufgabe (2.3) wurde in SPIRIT eingegeben und die Verteilung P^* berechnet. Die Randverteilungen von P^* auf allen Akteuren zeigen wir in Abbildung 3.2 und die zugehörigen Diffusionspotenziale und Rezeptionspotenziale in Abbildung 3.3.

Auffällig ist, dass auf den SZK $\{1, 6, 8, 9, 13, 17\}$, $\{7, 12\}$, $\{4, 5\}$ die Randverteilungen und damit die Diffusions- und Rezeptionspotenziale aller Akteure jeweils gleich sind. Es gibt zwar auch Akteure mit gleicher Randverteilung, die nicht in einer SZK liegen; sie werden

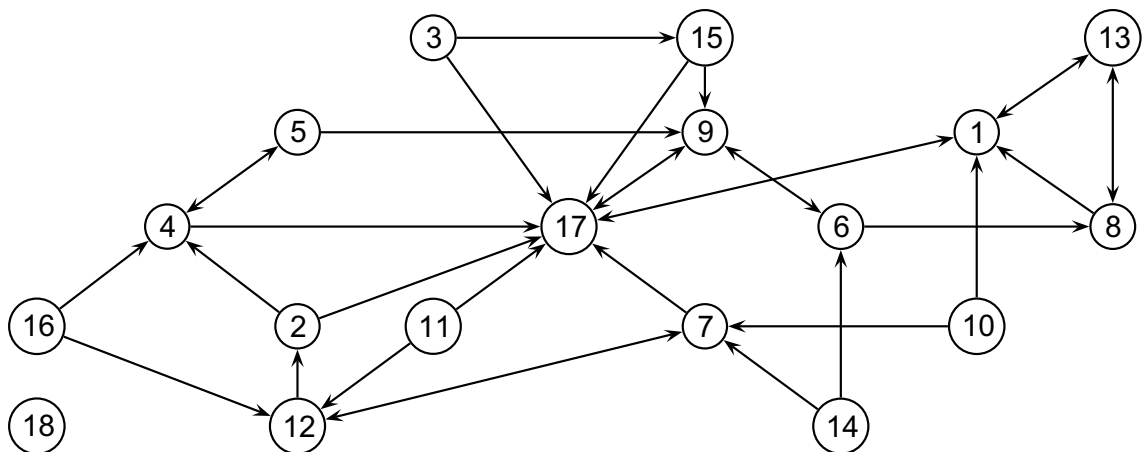


Abbildung 3.1: Modifiziertes Newcomb-Netz, Quelle: Rödder u. a. (2012)

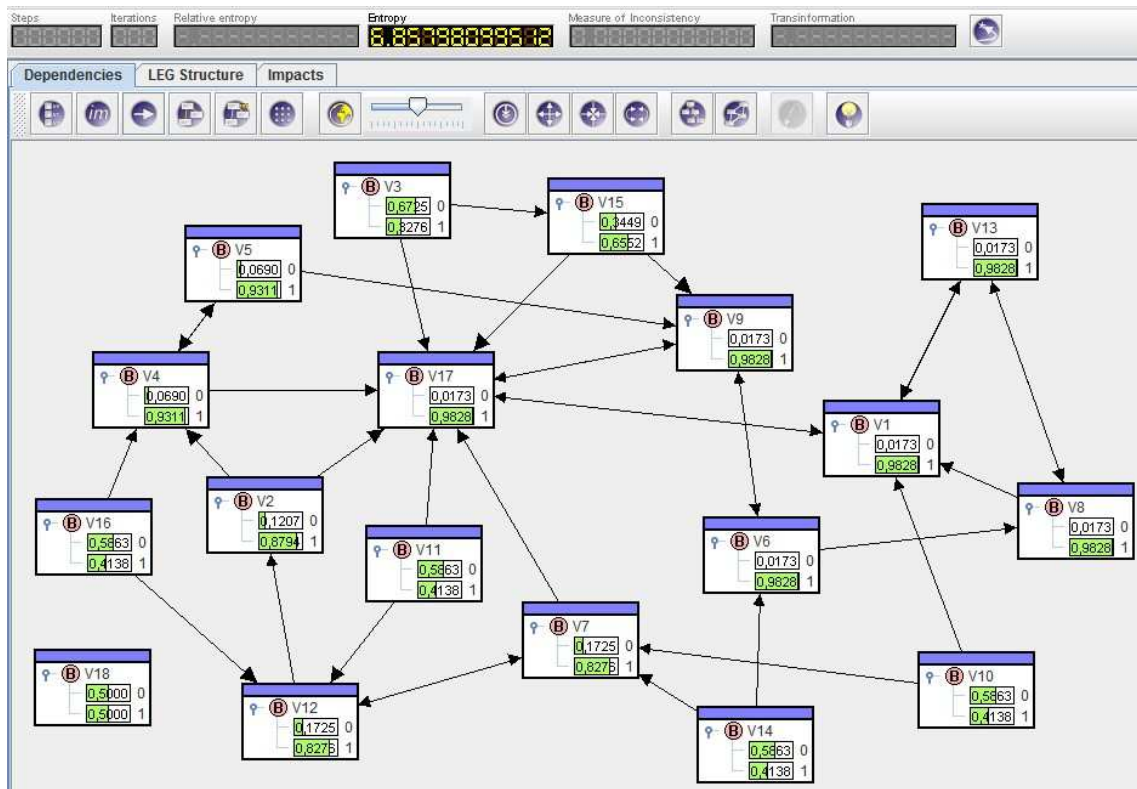


Abbildung 3.2: Randverteilungen P^* für alle Akteure, Quelle: Rödder u. a. (2012)

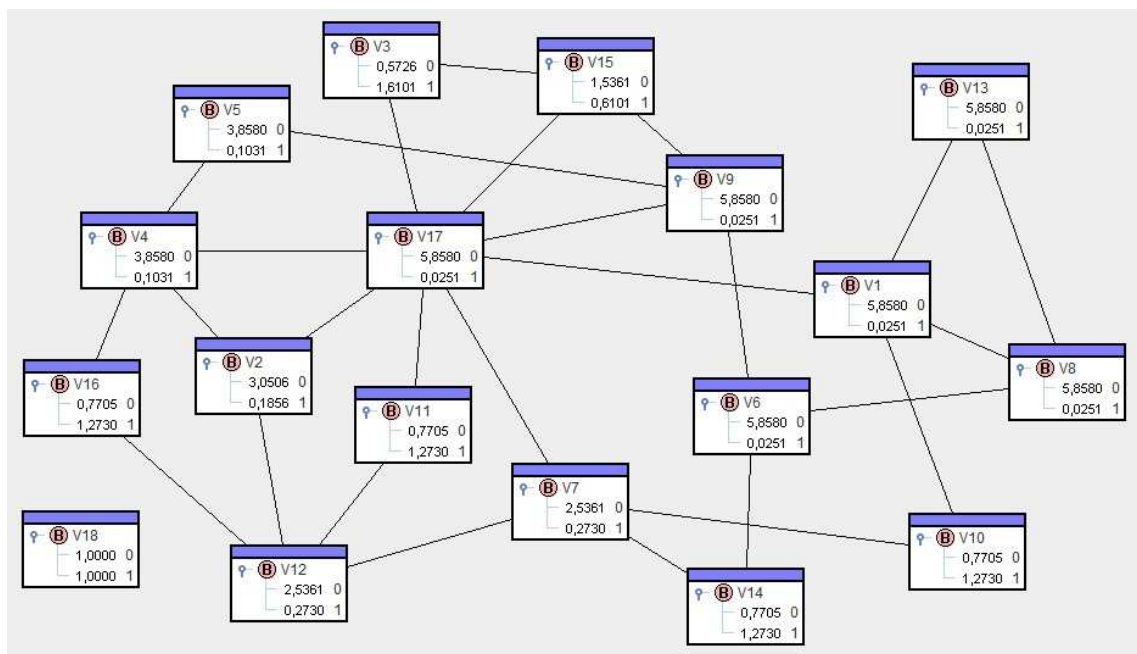


Abbildung 3.3: Diffusions- und Rezeptionspotenziale der Akteure, Quelle: [RöBrKu]

hier jedoch nicht weiter betrachtet. Siehe hierzu die Diskussion in Rödder u. a. (2012). Wir formulieren nun einen Satz, der die wesentlichen informationstheoretischen Charakterisierungen von SZK in einem SN zusammenfasst. Zwecks einfacher Schreibweise notieren wir ein Konditional der Form $V_l = 1 \mid V_k = 1[1.0]$ ab jetzt in der Form $l \mid k$.

Satz 1. Es sei ein SN mit der Konditionalstruktur $l_1 | k_1, \dots, l_I | k_I$ gegeben, es sei ferner $SZK \subset \{1, \dots, L\}$ die Indexmenge einer starken Zusammenhangskomponente $V_{SZK} \subset V$.

- i) Die Randverteilungen auf allen Akteuren $l \in SZK$ sind gleich.
- ii) Ist das gesamte SN eine SZK, sind die Randverteilungen auf allen Akteuren gleich und nehmen die Werte $P^*(V_l = 1) = P^*(V_l = 0) = \frac{1}{2}$ für alle $l \in SZK$ an.
- iii) Unter den Bedingungen wie in ii) ist $R(V_{SZK}; P^*, P^0) = L - 1$ und maximal für alle möglichen Konditionalstrukturen.

Beweis.

Zu i) Es sei $p_{1\dots 1}^* = p$ die Wahrscheinlichkeit des Zustands $1\dots 1$ von V_{SZK} . Für ein beliebiges $l \in SZK$ betrachte $p_{v_l=1, v_j \neq l}^*$ für alle $j \in SZK$. Ist $v_j = 0$ für mindestens ein j , gilt wegen $v_l = 1$: $p_{v_l=1, v_j \neq l}^* = 0.0$. Damit ist $p_{v_l=1}^* = p_{1\dots 1}^* + \sum_{v_j=0, \text{mind. ein } j} p_{v_l=1, v_j \neq l}^* = p + 0.0$. Da $l \in SZK$ beliebig gewählt wurde, hat man $p_{v_l=1}^* = p$ und $p_{v_l=0}^* = 1 - p$ für alle $l \in SZK$.

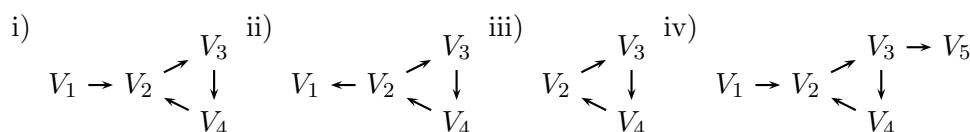
Zu ii) Wegen i) gilt mit $V_{SZK} = V$: Nur die Zustände $1\dots 1$ und $0\dots 0$ von V liefern positive Wahrscheinlichkeiten p und $1 - p$ und sie sind unter maximaler Entropie $p_{1\dots 1}^* = p_{0\dots 0}^* = \frac{1}{2}$.

Zu iii) Für die Verteilung P^* wie in ii) gilt $H(V; P^*) = 1$. Damit hat man $R(V; P^*, P^0) = \log_2 2^L - H(V; P^*) = L - 1$. Es bleibt zu zeigen: Für ein nicht stark zusammenhängendes Netz ist $R(V; P^*, P^0) < L - 1$. In diesem Fall nämlich haben auch andere Netzzustände als $1\dots 1$ und $0\dots 0$ eine positive Wahrscheinlichkeit. Ist die Zahl dieser Zustände m , gilt $H(V; P^*) = m[-\frac{1}{m} \log_2 \frac{1}{m}] > 1$ und somit $R(V; P^*, P^0) = \log_2 2^L - H(V; P^*) = L - (-\log_2 \frac{1}{m}) < L - 1$. ■

Teile des Satzes 1 wurden bereits in Rödder u. a. (2012) gezeigt, werden hier jedoch in einen erweiterten Kontext gestellt. Akteure in SZK sind informationstheoretisch nicht unterscheidbar, sie haben die gleichen Diffusions- und Rezeptionspotenziale. Im Übrigen ist die Konditionalstruktur eines SN dann ideal, wenn sie es zu einer einzigen SZK macht.

Im folgenden Abschnitt wollen wir die Frage untersuchen, ob SZK ohne größeren Wissensverlust zu einem Superknoten zusammengefasst werden können und welche informationstheoretischen Konsequenzen eine solche Knotenkompression hat. Zuvor jedoch noch eine Illustration der Ergebnisse des Satzes 1.

Beispiel 3. Für die folgenden Konditionalstrukturen geben wir die Verteilungen P^* an und erläutern:



Zu i) Die Lösung der Aufgabe (2.3) liefert

| V_1 | V_2 | V_3 | V_4 | P^* |
|-------|-------|-------|-------|---------------|
| 1 | 1 | 1 | 1 | $\frac{1}{3}$ |
| 1 | 1 | 1 | 0 | 0.0 |
| 1 | 1 | 0 | 1 | 0.0 |
| 1 | 1 | 0 | 0 | 0.0 |
| 1 | 0 | 1 | 1 | 0.0 |
| 1 | 0 | 1 | 0 | 0.0 |
| 1 | 0 | 0 | 1 | 0.0 |
| 1 | 0 | 0 | 0 | 0.0 |
| 0 | 1 | 1 | 1 | $\frac{1}{3}$ |
| 0 | 1 | 1 | 0 | 0.0 |
| 0 | 1 | 0 | 1 | 0.0 |
| 0 | 1 | 0 | 0 | 0.0 |
| 0 | 0 | 1 | 1 | 0.0 |
| 0 | 0 | 1 | 0 | 0.0 |
| 0 | 0 | 0 | 1 | 0.0 |
| 0 | 0 | 0 | 0 | $\frac{1}{3}$ |

Für die Randverteilungen hat man $P^*(V_1 = 1) = \frac{1}{3}$ und $P^*(V_2 = 1) = P^*(V_3 = 1) = P^*(V_4 = 1) = \frac{2}{3}$.

Zu ii) Die Lösung der Aufgabe (2.3) liefert

| V_1 | V_2 | V_3 | V_4 | P^* |
|-------|-------|-------|-------|---------------|
| 1 | 1 | 1 | 1 | $\frac{1}{3}$ |
| 1 | 1 | 1 | 0 | 0.0 |
| 1 | 1 | 0 | 1 | 0.0 |
| 1 | 1 | 0 | 0 | 0.0 |
| 1 | 0 | 1 | 1 | 0.0 |
| 1 | 0 | 1 | 0 | 0.0 |
| 1 | 0 | 0 | 1 | 0.0 |
| 1 | 0 | 0 | 0 | $\frac{1}{3}$ |
| 0 | 1 | 1 | 1 | 0.0 |
| 0 | 1 | 1 | 0 | 0.0 |
| 0 | 1 | 0 | 1 | 0.0 |
| 0 | 1 | 0 | 0 | 0.0 |
| 0 | 0 | 1 | 1 | 0.0 |
| 0 | 0 | 1 | 0 | 0.0 |
| 0 | 0 | 0 | 1 | 0.0 |
| 0 | 0 | 0 | 0 | $\frac{1}{3}$ |

Für die Randverteilungen hat man $P^*(V_1 = 1) = \frac{2}{3}$ und $P^*(V_2 = 1) = P^*(V_3 = 1) = P^*(V_4 = 1) = \frac{1}{3}$.

Zu iii) Die Lösung der Aufgabe (2.3) liefert

| V_2 | V_3 | V_4 | P^* |
|-------|-------|-------|---------------|
| 1 | 1 | 1 | $\frac{1}{2}$ |
| 1 | 1 | 0 | 0.0 |
| 1 | 0 | 1 | 0.0 |
| 1 | 0 | 0 | 0.0 |
| 0 | 1 | 1 | 0.0 |
| 0 | 1 | 0 | 0.0 |
| 0 | 0 | 1 | 0.0 |
| 0 | 0 | 0 | $\frac{1}{2}$ |

Für die Randverteilungen hat man $P^*(V_2 = 1) = P^*(V_3 = 1) = P^*(V_4 = 1) = \frac{1}{2}$.

Zu iv) Ohne die Lösung der Aufgabe (2.3) explizit aufzuschreiben, geben wir die Randverteilungen an: $P^*(V_1 = 1) = \frac{1}{4}$ $P^*(V_2 = 1) = P^*(V_3 = 1) = P^*(V_4 = 1) = \frac{1}{2}$ und $P^*(V_5 = 1) = \frac{3}{4}$.

Die Verteilungen P^* in Beispiel 3 zeigen die erwünschten Effekte; Randwahrscheinlichkeiten in SZK sind gleich. Ansonsten beobachten wir – wie schon in den Beispielen 1 und 2 – kleinere Wahrscheinlichkeiten $P^*(V_l = 1)$ bei größerem Diffusions- und kleinere Wahrscheinlichkeiten $P^*(V_l = 0)$ bei größerem Rezeptionspotenzial.

3.2 Kompression von SZK zu Superknoten

In diesem Abschnitt bezeichne wiederum $SZK \subset \{1, \dots, L\}$ die Indexmenge einer starken Zusammenhangskomponente $V_{SZK} \subset V$, $k \in SZK$ einen beliebigen Akteur aus dieser Komponente, $REST = \{1, \dots, L\} \setminus SZK$ die restlichen Knoten des Netzes, und es sei P^* die Lösung der Aufgabe (2.3).

Definition 1. (Knotenkompression)

Eine Knotenkompression nach V_{SZK} ist die Bildung des Netzes mit der Akteursmenge $V_\sigma \cup V_{REST}$, wobei der Akteur V_σ die Komponente V_{SZK} ersetzt. Die Verteilung $P^*|_{V_\sigma \cup V_{REST}}$ ist Lösung von (2.3) auf dem neuen Netz, d.h. unter Wegfall der Konditionale „in“ SZK und bei Übertragung der Konditionale zwischen $V_l, l \in SZK$ und $V_j, j \in REST$, auf V_σ und $V_j, j \in REST$.

Bemerkung: $P^*|_{V_\sigma \cup V_{REST}}$ – auch kurz $P^*|$ – ist identisch mit der Randverteilung von P^* auf $V_l \cup V_{REST}$ für ein beliebiges $l \in SZK$. Anstelle eines formalen Beweises führen wir Beispiel 3 fort.

Beispiel 4. (Fortsetzung von Beispiel 3)

Zu den Konditionalstrukturen i), ii), iii), iv) erhält man folgende reduzierte Graphen und die zugehörigen Randverteilungen.

Zu i) $V_1 \rightarrow V_\sigma$ ist der reduzierte Graph mit der Verteilung

| V_1 | V_σ | $P^* $ |
|-------|------------|---------------|
| 1 | 1 | $\frac{1}{3}$ |
| 1 | 0 | 0.0 |
| 0 | 1 | $\frac{1}{3}$ |
| 0 | 0 | $\frac{1}{3}$ |

Überprüfen Sie bitte, dass die Verteilung den Rand auf z.B. V_1, V_2 widerspiegelt.

Zu ii) $V_1 \leftarrow V_\sigma$ ist der reduzierte Graph mit der Verteilung

| V_1 | V_σ | $P^* $ |
|-------|------------|---------------|
| 1 | 1 | $\frac{1}{3}$ |
| 1 | 0 | $\frac{1}{3}$ |
| 0 | 1 | 0.0 |
| 0 | 0 | $\frac{1}{3}$ |

Überprüfen Sie bitte, dass die Verteilung den Rand auf z.B. V_1, V_2 widerspiegelt.

Zu iii) V_σ ist der reduzierte Graph mit der Verteilung

| V_σ | $\mathbf{P}^* \mathbf{I}$ |
|------------|---------------------------|
| 1 | $\frac{1}{2}$ |
| 0 | $\frac{1}{2}$ |

Wiederum stellt die Verteilung den Rand auf jedem $V_l \in V$ dar.

Zu iv) $V_1 \rightarrow V_\sigma \rightarrow V_5$ ist der reduzierte Graph mit der Verteilung

| V_1 | V_σ | V_5 | $\mathbf{P}^* \mathbf{I}$ |
|-------|------------|-------|---------------------------|
| 1 | 1 | 1 | $\frac{1}{4}$ |
| 1 | 1 | 0 | 0.0 |
| 1 | 0 | 1 | 0.0 |
| 1 | 0 | 0 | 0.0 |
| 0 | 1 | 1 | $\frac{1}{4}$ |
| 0 | 1 | 0 | 0.0 |
| 0 | 0 | 1 | $\frac{1}{4}$ |
| 0 | 0 | 0 | $\frac{1}{4}$ |

Auffällig in diesem Beispiel ist, dass der reduzierte Graph

- wie in i) die gleiche Verteilung wie in Beispiel 1 i),
- wie in iv) die gleiche Verteilung wie in Beispiel 1 ii) hat.

Ist die SZK einmal zum Superknoten komprimiert, ist er informationstheoretisch nicht mehr als solcher zu erkennen. Das reduzierte Netz ist einem strukturgleichen ohne Superknoten äquivalent. Das wird nun in einem Satz festgehalten.

Satz 2.

- Die Entropien im Netz sind vor und nach Knotenkompression gleich.
- Der Superknoten V_σ im reduzierten Netz hat das gleiche Diffusions- und Rezeptionspotenzial wie jeder Knoten der SZK vor Kompression.

Beweis.

Es sei $k \in SZK$ ein beliebig fester Knoten.

Zu i) Mit der in diesem Abschnitt eingeführten Symbolik sei $v = v_k v_{SZK \setminus \{k\}} v_{REST}$ ein beliebiger Zustand des Netzes und \mathbf{P}^* die Lösung von (2.3). Faktorisiere

$$\mathbf{P}^*(v) = \mathbf{P}^*(v_k) \cdot \mathbf{P}^*(v_{SZK \setminus \{k\}} | v_k) \cdot \mathbf{P}^*(v_{REST} | v_{SZK}).$$

Damit ergibt sich für die Entropie des Netzes

$$\begin{aligned} H(V; \mathbf{P}^*) &= - \sum_{\text{alle } v} \mathbf{P}^*(v) \log_2 \mathbf{P}^*(v) = - \sum_{\text{alle } v_k} \mathbf{P}^*(v_k) \log_2 \mathbf{P}^*(v_k) \\ &- \sum_{\text{alle } v_k} \mathbf{P}^*(v_k) \sum_{\text{alle } v_{SZK \setminus \{k\}}} \mathbf{P}^*(v_{SZK \setminus \{k\}} | v_k) \log_2 \mathbf{P}^*(v_{SZK \setminus \{k\}} | v_k) \\ &- \sum_{\text{alle } v_{SZK}} \mathbf{P}^*(v_{SZK}) \sum_{\text{alle } v_{REST}} \mathbf{P}^*(v_{REST} | v_{SZK}) \log_2 \mathbf{P}^*(v_{REST} | v_{SZK}) \end{aligned}$$

Nun ist aber der mittlere Term – also die bedingte Entropie gegeben V_k – gleich 0.0, da alle auftretenden bedingten Wahrscheinlichkeiten 1.0 oder 0.0 sind. Damit hat man $H(V; P^*) = H(V_k; P^*) + \underbrace{H(V_{SZK} | V_k; P^*)}_{=0.0} + H(V_{REST} | V_{SZK}; P^*)$.

Nun sei das reduzierte Netz betrachtet. O.B.d.A. sei V_k der Superknoten V_σ und \bar{V} die Akteursmenge. $\bar{v} = v_k v_{REST}$ sei wiederum ein beliebiger Zustand des Netzes und $P^* I_{\{k\} \cup REST}$ oder einfach $P^* I$ die Randverteilung von P^* . Mit der Faktorisierung $P^* I(\bar{v}) = P^* I(v_k) P^* I(v_{REST} | v_k)$ erhält man jetzt

$$\begin{aligned} H(\bar{V}; P^* I) &= H(V_k; P^* I) + H(V_{REST} | V_k; P^* I) \\ &= H(V_k; P^*) + H(V_{REST} | V_k; P^*) \end{aligned}$$

Zu ii) Die Aussage ist eine unmittelbare Folge aus der Tatsache, dass die Randverteilung auf V_σ identisch mit der jedes Akteurs in SZK ist. ■

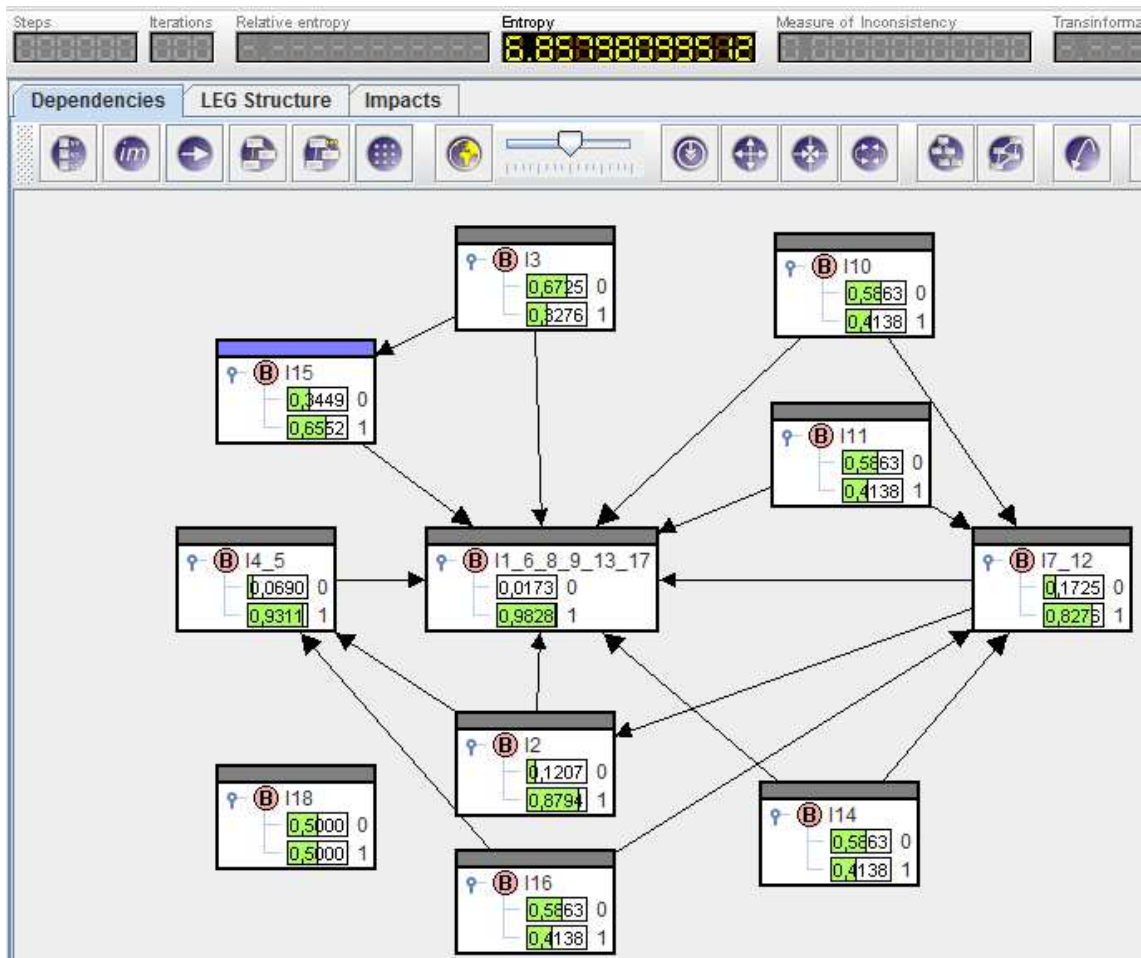


Abbildung 3.4: Komprimiertes Newcomb-Netzwerk

Zusammenfassend lässt sich also feststellen:

- Die Entropien des ursprünglichen und des reduzierten Netzes sind gleich.
- Der Superknoten hat die gleiche Randverteilung wie jedes $V_k \in V_{SZK}$ im ursprünglichen Netz.
- Damit hat der Superknoten im reduzierten Netz auch das gleiche Diffusions- und Rezeptionspotenzial wie jedes $V_k \in V_{SZK}$ im ursprünglichen Netz.

Hat ein Netz zahlreiche SZK, kann der Reduktionsvorgang solange wiederholt werden, bis alle SZK zu Superknoten komprimiert sind; die Entropie bleibt dabei stets gleich. Informationstheoretisch entstehen also stets äquivalente Netze. Die Aussagen (2.5) und (2.6) aus Abschnitt 2.2 sind auch im reduzierten Netz mit P^* statt P^* gültig. Informationstheoretisch äquivalent bedeutet natürlich(!) nicht graphentheoretisch äquivalent. Inzidenzen und Adjazenzen ändern sich in dem Maße, wie ursprüngliche Akteure in SZK aus dem Netz entfernt werden und in die Superknoten „einfließen“. Aufgrund der Transitivität der Konditionale – siehe wiederum Abschnitt 2.2 – sind alle Akteure in einer SZK untereinander erreichbar, sie alle erhalten also das (im-)materielle Gut oder die Botschaft, falls es nur ein Akteur erhält. Kompression von SZK zu Superknoten ist ein probates Mittel zur Vereinfachung der Netzstruktur zwischen den Akteuren. Zur Illustration der Aussage greifen wir wieder das Newcomb-Netz auf und zeigen in Abbildung 3.4 den zu Abbildung 3.3 reduzierten Graphen mit den Superknoten 1_6_8_9_13_17, 7_12, 4_5. In jedem Superknoten sind die Randwahrscheinlichkeiten erkennbar; sie entsprechen denen der Akteure in den SZK des nicht komprimierten Netzes.

Kapitel 4

Netzreduktion durch Kompression

Für viele Anwendungen von SN trägt eine Reduktion des Netzes zur Transparenz bei. Gruppierungen von Akteuren in SZK werden erkennbar und die Abhängigkeiten zwischen den Gruppen ebenfalls. In diesem Kapitel demonstrieren wir das Reduktionspotenzial in größeren Netzen an zwei Beispielen, je einem Zufallsnetz mit 750 und 672 Akteuren.

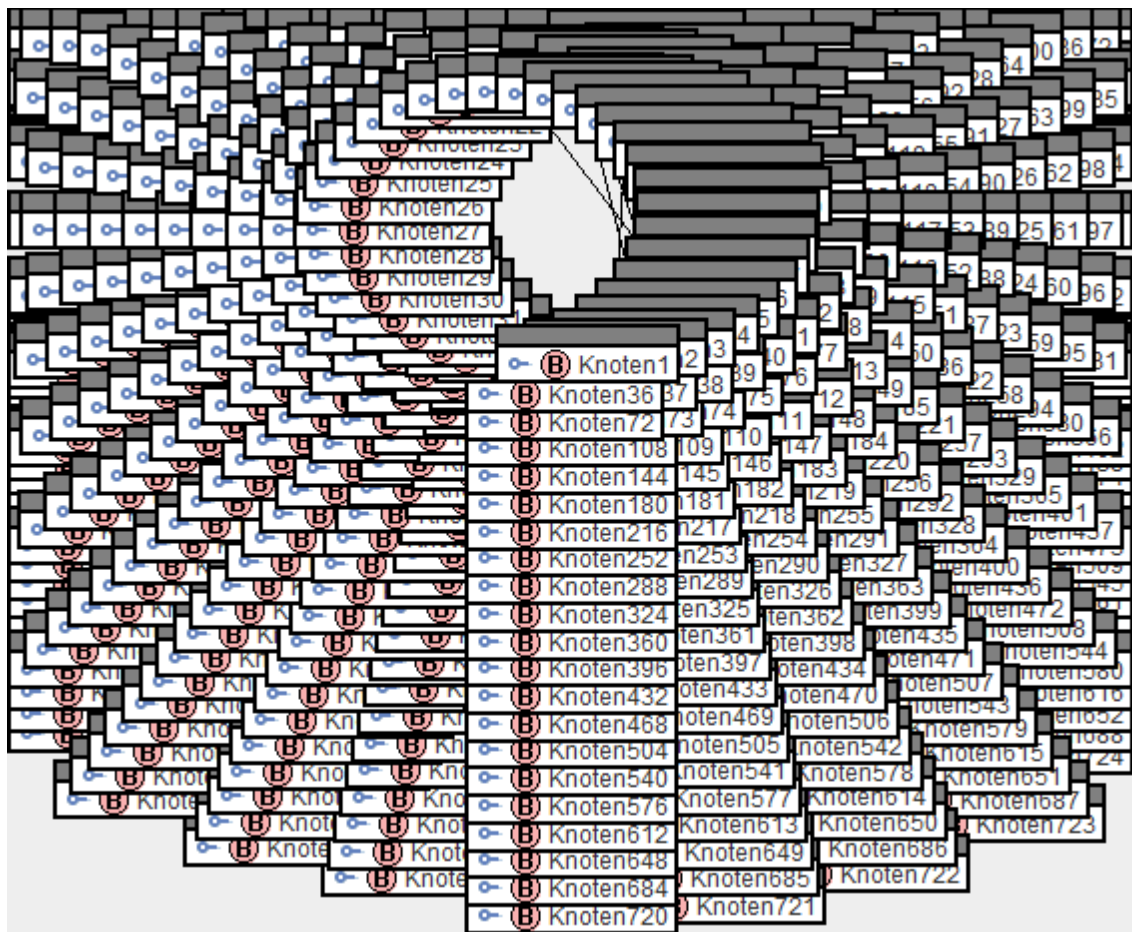


Abbildung 4.1: Ausschnitt eines Zufallsnetzes mit 750 Knoten

indem eine gewünschte Anzahl an Akteuren zufällig ausgewählt und anschließend zu einem geschlossenen Kantenzug verbunden wird. Dabei sollen jedoch nur die Akteure miteinander verbunden werden, die noch nicht Element einer SZK sind. Das Vorgehen wird beliebig oft wiederholt.

Anschließend können die Konditionale $V_l | V_k[1.0]$ der Shell SPIRIT übergeben werden.

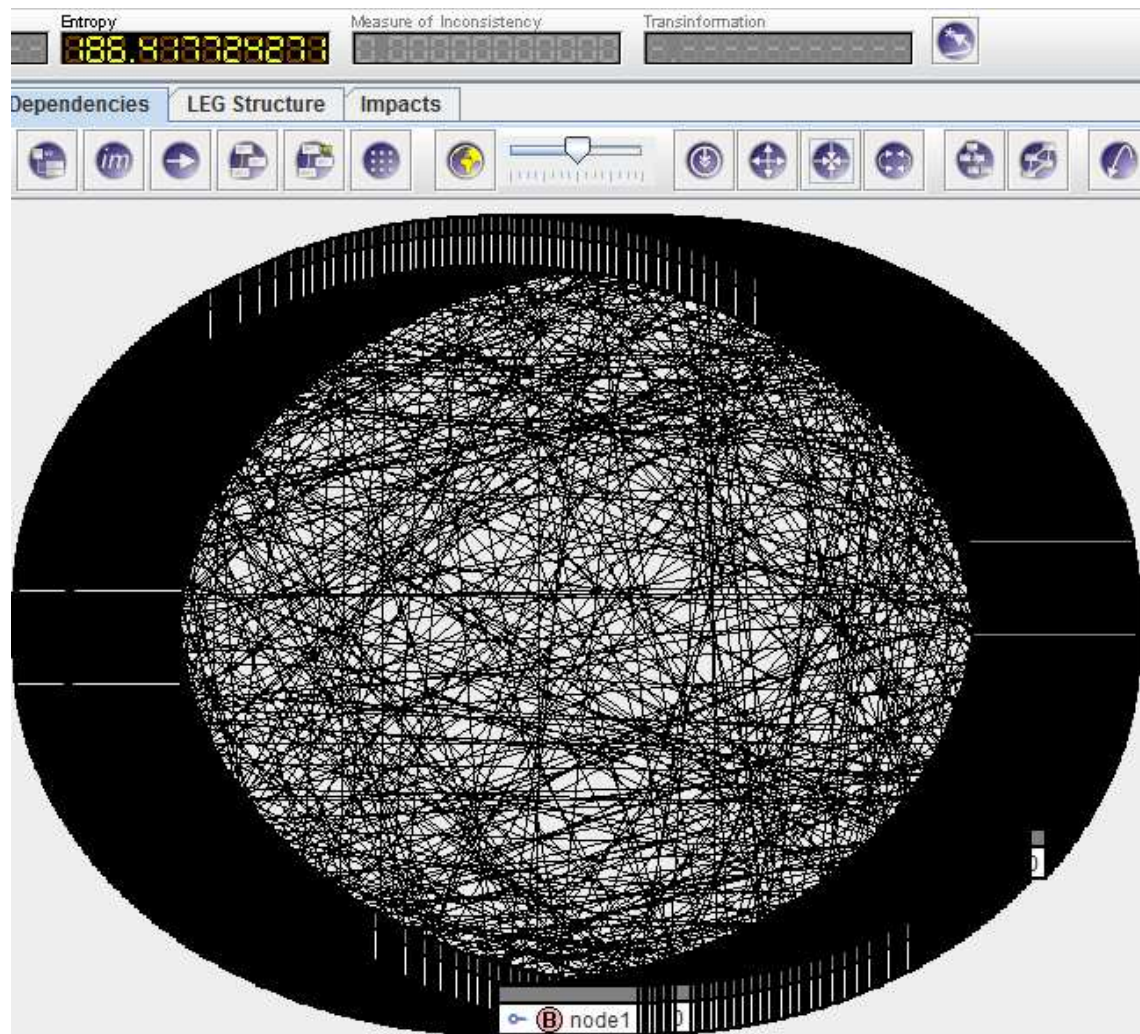


Abbildung 4.3: Synthetisches Netz mit 672 Knoten

Beispiel 6. Ein so erzeugtes Netz zeigt Abbildung 4.3. Mit Parametern $L = 1500$ und $q = 0.000175$ wurde ein Netzwerk mit insgesamt 672 Akteuren erzeugt. Zusätzlich wurde in 1000 Iterationen versucht, SZK mit einer Größe von 6 Mitgliedern zu erzeugen. Das resultierende Netzwerk enthält 141 SZK mit der in Tabelle 4.1 gezeigten Größenverteilung. Nach Eingabe in SPIRIT weist die Shell eine Entropie von 186.42 [bit] bei 1703 Konditionalen aus. Nach dem oben beschriebenen Reduktionsprozess enthält das Netz noch 388

| | | | | | | | | |
|------------|----|----|----|----|---|----|----|----|
| Mitglieder | 2 | 3 | 4 | 5 | 6 | 12 | 17 | 29 |
| Anzahl | 69 | 26 | 23 | 11 | 9 | 1 | 1 | 1 |

Tabelle 4.1: Größenverteilung der SZK des 672er Netzwerks

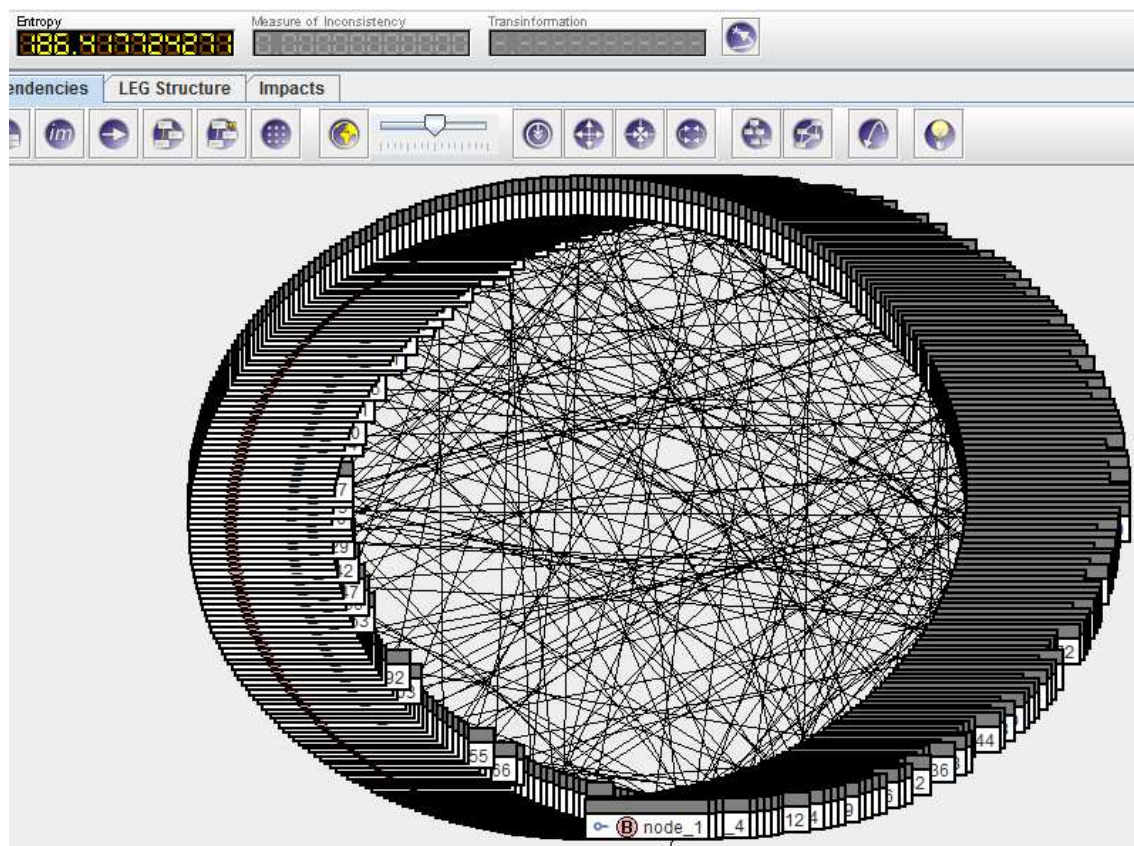


Abbildung 4.4: Komprimiertes synthetisches Netz mit 388 Knoten

(Super-)Knoten bei natürlich gleicher Entropie. Die Anzahl der Konditionale reduziert sich auf 535. Abbildung 4.4 illustriert die erhebliche Reduktion.

Ein weiteres Beispiel, an dem auch gleichzeitig wichtige Aussagen dieses Berichts demonstriert werden können, stammt von Opsahl u. Panzarasa (2009) und enthält 1899 Knoten mit 20296 Konditionalen. Wie Abbildung 4.5 anschaulich zeigt, handelt es sich dabei um ein eher dichtes Netz. Trotz ausgeklügelter Berechnungsverfahren (vgl. Meyer, 1998) stößt SPIRIT hier sowohl hinsichtlich des Aufbaus der Verteilung als auch des grafischen Darstellungsvermögens an Grenzen. Die Darlegung der Gründe hierfür sprengen den Rahmen dieses Arbeitsberichts. Stattdessen nutzen wir Satz 2 aus und führen unsere Analyse am komprimierten Netz durch. Vorher werfen wir noch einen genaueren Blick auf die Strukturen des Originalnetzwerks. Dort finden sich 5 bilaterale SZK (SZK 2 - SZK 6) sowie eine große SZK (SZK 1) mit 1294 Mitgliedern. 595 Knoten sind in keiner SZK (Peripherie). Von den Mitgliedern der SZK 1 werden 20170 ausgehende und 19117 eingehende Kontakte unterhalten. Davon dienen jeweils 19026 der Kommunikationen innerhalb der SZK. Lediglich 1144 der ausgehenden Pfeile haben Destinationen außerhalb der Komponente, eine verschwindende Anzahl von 91 Beziehungen finden ihr Ziel dorthin von außerhalb. Allein diese Zahlen lassen erwarten, dass die Kompression eindrucksvolle Effekte erzeugen wird und diese Erwartung wird nicht enttäuscht: Das komprimierte Netz enthält nur noch 601 Knoten, davon natürlich 6 Superknoten, und 614 Konditionale. Abbildung 4.6 ist zum einen die umfangreiche Komplexitätsreduktion anzusehen, zum anderen zeigt sich, dass sich das Netzwerk um den Superknoten 1 herum gruppiert.

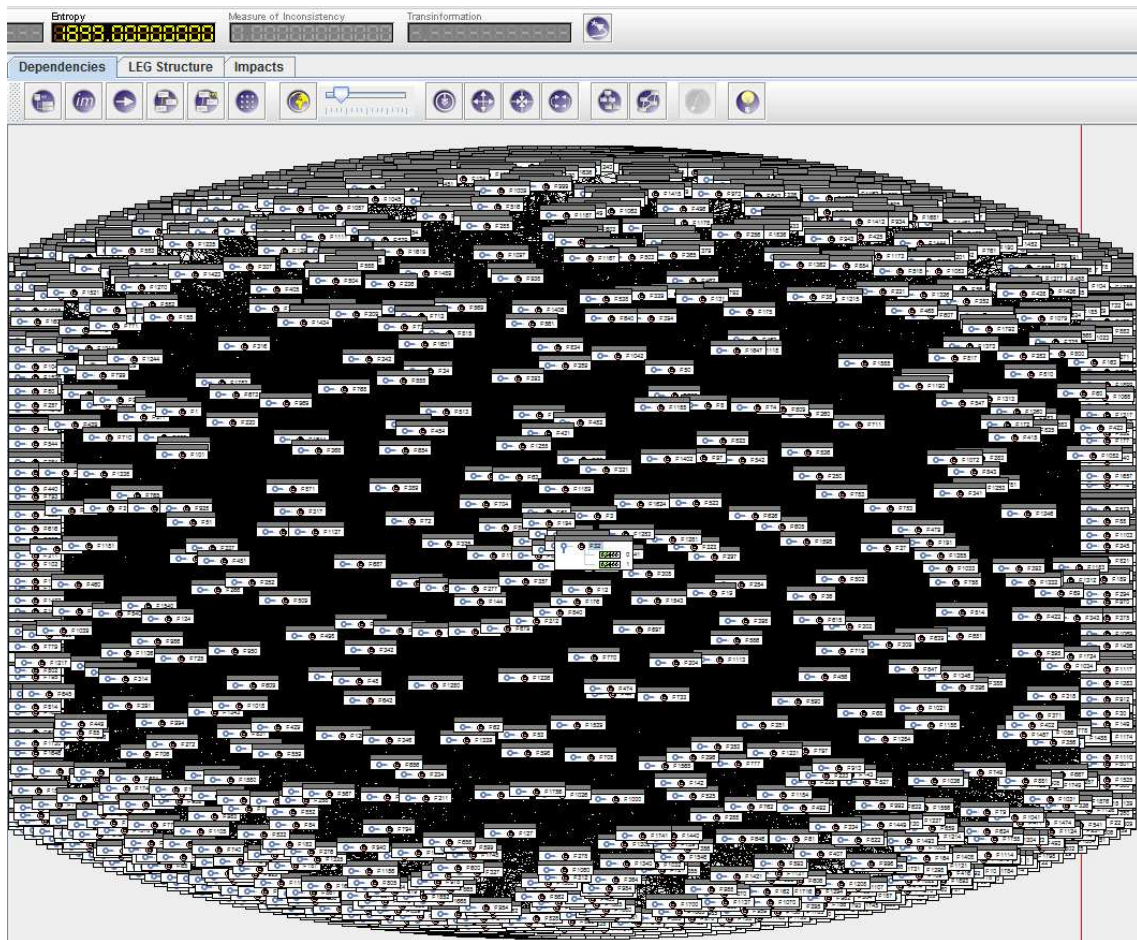


Abbildung 4.5: Facebook-like network, in Anlehnung an Opsahl u. Panzarasa (2009)

Das Diffusionspotenzial des Superknotens 1 ist immens. Der Abbildung 4.6 ist es leider nicht zu entnehmen, daher geben wir den Wert hier an: etwa 519 [bit]. Das entspricht einer Wahrscheinlichkeit von ca. 10^{-156} für die Variablenausprägung 1, das Rezeptionspotenzial ist dagegen vernachlässigbar gering mit der Gegenwahrscheinlichkeit von $1 - 10^{-156}$. Die extreme Ausprägung der Potenziale lässt sich auch hier mittels der Struktur des reduzierten Netzwerks erläutern. Von den insgesamt 614 Konditionalen beschreiben 555 die ausgehende und nur 34 die eingehende Kommunikation des Superknotens 1. Damit nimmt

Tabelle 4.2: Zentralitäten vor und nach Kompression Opsahls Netzwerk

| Knoten-Nr. | Opsahls Netz | | komprimiertes Opsahl-Netz | |
|------------|--------------|---------------|---------------------------|-----------|
| | c_D^+ | Netzwergruppe | Knoten-Nr. | c_D^+ |
| 32 | 0,0721812 | SZK 1 | Super 1 | 0,1516667 |
| 784 | 0,0136986 | Peripherie | 784 | 0,0433333 |
| 10 | 0,0010537 | SZK 2 | Super 10 | 0,0016667 |
| 433 | 0,0010537 | SZK 5 | Super 433 | 0,0033333 |
| 20 | 0,0031612 | SZK 3 | Super 20 | 0,0083333 |
| 187 | 0,0005269 | SZK 4 | Super 187 | 0 |
| 1797 | 0,0005269 | SZK 6 | Super 1797 | 0 |

Superknoten 1 ganz offensichtlich eine zentrale Position im Netzwerk ein – ruft man sich den Ursprung der Daten in Erinnerung, scheint es nicht falsch, diese als „information

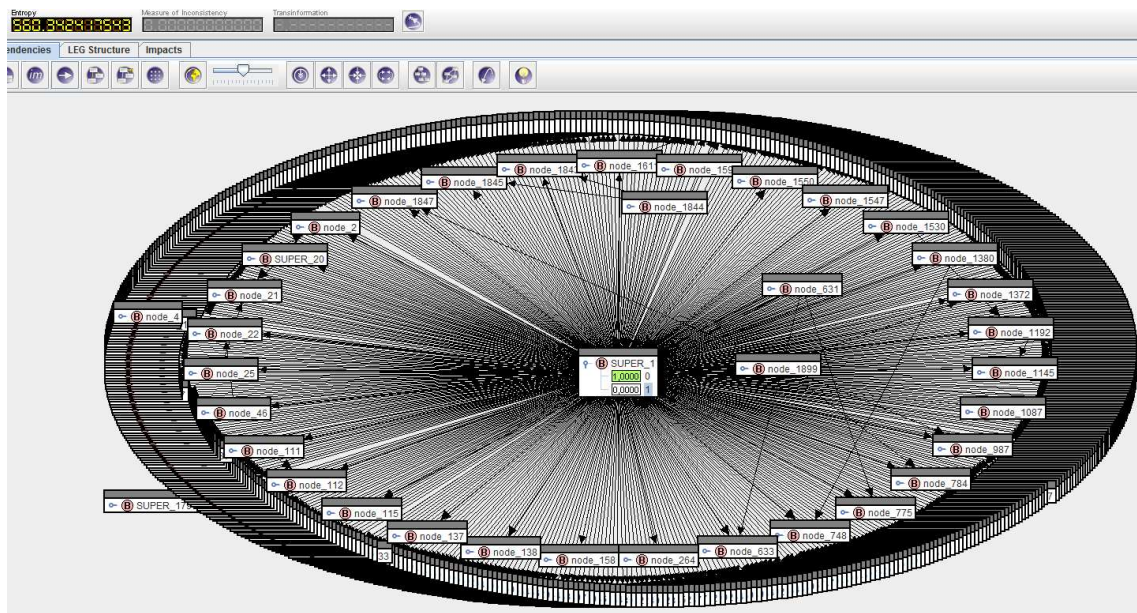


Abbildung 4.6: Reduziertes Netz zu Abbildung 4.5

hub“ (vgl. bspw. Kleinberg, 1999) zu bezeichnen.

Tabelle 4.2 zeigt einen Vergleich der out degree-Zentralitäten (c_D^+) (vgl. Freeman, 1979) ausgewählter Akteure der einzelnen Netzwerkgruppen vor und nach Kompression des Netzwerks.

Die Berechnungen wurden mit Scilab ausgeführt. Nicht überraschend ist der „zentralste“ Akteur (32) ein Vertreter der SZK 1. Den höchsten Zentralitätswert außerhalb von SZK 1 nimmt der Akteur 784 ein; er gehört keiner SZK an. Die Vertreter der anderen SZK folgen mit deutlichem Abstand. Wie in Abschnitt 3.2 erwähnt, ändern sich die Zentralitäten natürlich mit der Kompression. Akteur 32 wird nun von dem Superknoten 1 repräsentiert und weist eine vergleichsweise höhere Zentralität aus. Auch der Peripherieknoten behält seine hohe Zentralität bei. Die Ergebnisse der Knoten Super 10, Super 433 sowie Super 20 sind unspektakulär. Für die SZK 187 bzw. 1797 sinkt die Zentralität hingegen auf Null. Für die jeweiligen Superknoten kann festgehalten werden, dass ihre Zentralität lediglich auf dem bilateralen Verhältnis beruhte. Beide SZK bzw. Superknoten sind lediglich „Informationsempfänger“. Dieser Zusammenhang ist rein graphentheoretischer Natur. Was bei der graphentheoretischen Betrachtung nicht offensichtlich wird, ist, dass sich das Diffusions- und Rezeptionspotenzial auch bei Kompression nicht ändert. Etwas unsauber ausgedrückt, ist ihre informationstheoretische Position invariant gegenüber der Netzreduktion.

Kapitel 5

Zusammenfassung und Ausblick

Im Diskussionsbeitrag Nr. 481 der Fakultät für Wirtschaftswissenschaften an der Fern-Universität in Hagen wurde eine neue Form der Analyse von Interaktionen in Netzwerken vorgestellt. Dazu prägt man einem gerichteten Netzwerk eine Konditionalstruktur und in Folge eine solche Wahrscheinlichkeitsverteilung auf, deren Entropie minimal von der der Gleichverteilung abweicht. Das damit intendierte Vorsichtsprinzip besagt: Wähle eine solche probabilistische Abhängigkeitsstruktur, die kein nicht mitgeteiltes Wissen über das Netz beinhaltet.

Für solche informationstheoretisch aufbereiteten Netze konnten dann neue Analyseinstrumente vorgestellt werden:

- ein Maß für die (Struktur-)Dichte des Netzes,
- Maße für das Diffusions- und Rezeptionspotenzial aller Akteure,
- ein Maß für die Einbindung jedes Akteurs ins Netz

etc.

Multiple Netze – d.h. Netze mit mehreren Interaktionsarten – konnten ebenso modelliert werden, wie die Trennung von Akteur und dem Interaktionssubstrat. Alle Netze sind in der Expertensystem-Shell SPIRIT abbildbar, und die genannten Maße werden von dieser Shell bereitgestellt.

Bereits bei diesen Untersuchungen fiel auf, dass Akteure in sogenannten starken Zusammenhangskomponenten des Netzes informationstheoretisch nicht unterscheidbar sind; sie alle haben gleiches Diffusions- und Rezeptionspotenzial. Hier setzt der aktuelle Diskussionsbeitrag an.

Zunächst wird die Darstellung der Wissensakquisition vereinfacht und auf die Lösung der zu stellenden Fragestellungen zugeschnitten. Dann erfolgt eine informationstheoretische Charakterisierung von solchen Zusammenhangskomponenten und ihre Verdichtung zu sogenannten Superknoten. Superknoten ersetzen diese Komponenten, wahren dabei jedoch die gesamte konditionale Einbindung ins Netz. Es kann dann gezeigt werden, dass die Verdichtung von – auch mehreren – Zusammenhangskomponenten den informationstheoretischen Charakter des Netzes nicht ändert. Ein wenig lax gesagt: Die verbleibende Unsicherheit=Entropie nach Erlernen der Struktur im *reduzierten* Netz ist der verbleibenden Unsicherheit (nach Erlernen der Struktur) im *ursprünglichen* Netz gleich.

Der Vorteil solcher Reduktionen liegt auf der Hand:

- Sie filtern Zusammenhangskomponenten aus dem Netz heraus,
- sie machen die Abhängigkeiten zwischen Akteursgruppen transparent.

Ein kleines (18 Akteure mit 34 Konditionalen) und zwei größere (750 Akteure mit 756 Konditionalen, 672 Akteure und 1703 Konditionale) werden untersucht, reduziert und ihre informationstheoretischen Charakteristika erarbeitet. Die Reduktionseffekte sind beeindruckend. Es schließen sich einige Bemerkungen über das Netzwerk von Opsahl u. Panzarasa (2009) an. Dieses „facebook-like“ Netz hat 1899 Knoten und 20296 Konditionale. In dieser Form entzieht es sich der Berechenbarkeit durch SPIRIT (Rödder, 2009), nicht aber das reduzierte Netz. In der reduzierten Form lassen sich informationstheoretische Eigenschaften von Superknoten ableiten, die Schlüsse über das Originalnetz ermöglichen und über rein graphentheoretische Analysen hinausgehen.

Wie bereits in Diskussionsbeitrag Nr. 481 erwähnt, sind die bisher untersuchten Netze ein Spezialfall insofern, als alle Konditionale sicher sind. Die Frage der Berechnung aller informationstheoretischen Indizes wie Netzdichte, Diffusion, Rezeption und Einbindung für den Fall, dass die Konditionale nicht sicher, sondern mit einer von 1.0 verschiedenen Wahrscheinlichkeit gelten, ist Gegenstand zukünftiger Untersuchungen.

Literaturverzeichnis

- [Barnes 1969] BARNES, John A.: Networks and political process. In: MITCHELL, James C. (Hrsg.): *(Social networks in urban situations)*. Manchester : Manchester University Press, 1969, S. 1–50
- [Borgatti 2005] BORGATTI, S.: Centrality and network flow. In: *Social Networks* 27 (2005), January, Nr. 1, S. 55–71
- [Freeman 1979] FREEMAN, Linton C.: Centrality in social networks conceptual clarification. In: *Social Networks* 1 (1979), Nr. 3, S. 215–239
- [Kleinberg 1999] KLEINBERG, Jon M.: Authoritative sources in a hyperlinked environment. In: *Journal of the ACM (JACM)* 46 (1999), Nr. 5, S. 604–632
- [Meyer 1998] MEYER, Carl H.: *Korrektcs SchlieÙen bei unvollstcndiger Information*, FernUniversitcet in Hagen, Diss., 1998
- [Newcomb 1961] NEWCOMB, Theodore M.: *The acquaintance process*. Holt, Rinehart and Winston, 1961 (Report (Western Behavioral Sciences Institute))
- [Newman 2010] NEWMAN, Mark E. J.: *Networks: An Introduction*. OUP Oxford, 2010
- [Opsahl u. Panzarasa 2009] OPSAHL, Tore ; PANZARASA, Pietro: Clustering in weighted networks. In: *Social Networks* 31 (2009), Nr. 2, S. 155 – 163
- [Rödder u. a. 2012] RÖDDER, Wilhelm ; BRENNER, Dominic ; KULMANN, Friedhelm: Entropiebasierte Analyse von Interaktionen in Sozialen Netzwerken. In: *Diskussionsbeitrag der Fakultät für Wirtschaftswissenschaft*, FernUniversitcet in Hagen, 2012 (481)
- [Rödder u. Kern-Isberner 1997] RÖDDER, Wilhelm ; KERN-ISBERNER, Gabriele: Léa Sombé und entropie-optimale Informationsverarbeitung mit der Expertensystem-Shell SPIRIT. In: *Operations-Research-Spektrum* 19 (1997), Nr. 1, S. 41–46
- [Rödder 2009] RÖDDER, Wilhelm: *SPIRITshell*. 3.8.1.3.Extended. Hagen: FernUniversitcet in Hagen, 2009. – <http://xspirit.de/> (19.09.2013)
- [Scilab Enterprises 2012] SCILAB ENTERPRISES: *Scilab: Free and Open Source software for numerical computation*. 5.4.0. Orsay, France: Scilab Enterprises, 2012. – <http://www.scilab.org> (19.09.2013)
- [Scott 1991] SCOTT, John: *Social network Analysis*. London : Sage, 1991
- [Trappmann u. a. 2005] TRAPPMANN, Mark ; HUMMELL, Hans J. ; SODEUR, Wolfgang: *Strukturanalyse sozialer Netzwerke. Konzepte, Modelle, Methoden*. VS Verlag, 2005
- [Wasserman u. Faust 1994] WASSERMAN, Stanley ; FAUST, Katherine: *Social network analysis: methods and applications*. Cambridge University Press : Cambridge, 1994

Die Diskussionspapiere ab Nr. 183 (1992) bis heute, können Sie im Internet unter <http://www.fernuni-hagen.de/wirtschaftswissenschaft/forschung/beitraege.shtml> einsehen und zum Teil downloaden.

Ältere Diskussionspapiere selber erhalten Sie nur in den Bibliotheken.

| Nr | Jahr | Titel | Autor/en |
|-----|------|---|--|
| 420 | 2008 | Stockkeeping and controlling under game theoretic aspects | Fandel, Günter Trockel, Jan |
| 421 | 2008 | On Overdissipation of Rents in Contests with Endogenous Intrinsic Motivation | Schlepütz, Volker |
| 422 | 2008 | Maximum Entropy Inference for Mixed Continuous-Discrete Variables | Singer, Hermann |
| 423 | 2008 | Eine Heuristik für das mehrdimensionale Bin Packing Problem | Mack, Daniel Bortfeldt, Andreas |
| 424 | 2008 | Expected A Posteriori Estimation in Financial Applications | Mazzoni, Thomas |
| 425 | 2008 | A Genetic Algorithm for the Two-Dimensional Knapsack Problem with Rectangular Pieces | Bortfeldt, Andreas Winter, Tobias |
| 426 | 2008 | A Tree Search Algorithm for Solving the Container Loading Problem | Fanslau, Tobias Bortfeldt, Andreas |
| 427 | 2008 | Dynamic Effects of Offshoring | Stijepic, Denis Wagner, Helmut |
| 428 | 2008 | Der Einfluss von Kostenabweichungen auf das Nash-Gleichgewicht in einem nicht-kooperativen Disponenten-Controller-Spiel | Fandel, Günter Trockel, Jan |
| 429 | 2008 | Fast Analytic Option Valuation with GARCH | Mazzoni, Thomas |
| 430 | 2008 | Conditional Gauss-Hermite Filtering with Application to Volatility Estimation | Singer, Hermann |
| 431 | 2008 | Web 2.0 auf dem Prüfstand: Zur Bewertung von Internet-Unternehmen | Christian Maaß Gotthard Pietsch |
| 432 | 2008 | Zentralbank-Kommunikation und Finanzstabilität – Eine Bestandsaufnahme | Knütter, Rolf Mohr, Benjamin |
| 433 | 2008 | Globalization and Asset Prices: Which Trade-Offs Do Central Banks Face in Small Open Economies? | Knütter, Rolf Wagner, Helmut |
| 434 | 2008 | International Policy Coordination and Simple Monetary Policy Rules | Berger, Wolfram Wagner, Helmut |
| 435 | 2009 | Matchingprozesse auf beruflichen Teilarbeitsmärkten | Stops, Michael Mazzoni, Thomas |
| 436 | 2009 | Wayfindingprozesse in Parksituationen - eine empirische Analyse | Fließ, Sabine Tetzner, Stefan |
| 437 | 2009 | ENTROPY-DRIVEN PORTFOLIO SELECTION a downside and upside risk framework | Röder, Wilhelm Gartner, Ivan Ricardo Rudolph, Sandra |
| 438 | 2009 | Consulting Incentives in Contests | Schlepütz, Volker |

| | | | |
|-----|------|--|---|
| 439 | 2009 | A Genetic Algorithm for a Bi-Objective Winner-Determination Problem in a Transportation-Procurement Auction" | Buer, Tobias Pankratz, Giselher |
| 440 | 2009 | Parallel greedy algorithms for packing unequal spheres into a cuboidal strip or a cuboid | Kubach, Timo Bortfeldt, Andreas Tilli, Thomas Gehring, Hermann |
| 441 | 2009 | SEM modeling with singular moment matrices Part I: ML-Estimation of time series | Singer, Hermann |
| 442 | 2009 | SEM modeling with singular moment matrices Part II: ML-Estimation of sampled stochastic differential equations | Singer, Hermann |
| 443 | 2009 | Konsensuale Effizienzbewertung und -verbesserung – Untersuchungen mittels der Data Envelopment Analysis (DEA) | Röder, Wilhelm Reucher, Elmar |
| 444 | 2009 | Legal Uncertainty – Is Harmonization of Law the Right Answer? A Short Overview | Wagner, Helmut |
| 445 | 2009 | Fast Continuous-Discrete DAF-Filters | Mazzoni, Thomas |
| 446 | 2010 | Quantitative Evaluierung von Multi-Level Marketingsystemen | Lorenz, Marina Mazzoni, Thomas |
| 447 | 2010 | Quasi-Continuous Maximum Entropy Distribution Approximation with Kernel Density | Mazzoni, Thomas Reucher, Elmar |
| 448 | 2010 | Solving a Bi-Objective Winner Determination Problem in a Transportation Procurement Auction | Buer, Tobias Pankratz, Giselher |
| 449 | 2010 | Are Short Term Stock Asset Returns Predictable? An Extended Empirical Analysis | Mazzoni, Thomas |
| 450 | 2010 | Europäische Gesundheitssysteme im Vergleich – Effizienzmessungen von Akutkrankenhäusern mit DEA – | Reucher, Elmar Sartorius, Frank |
| 451 | 2010 | Patterns in Object-Oriented Analysis | Blaimer, Nicolas Bortfeldt, Andreas Pankratz, Giselher |
| 452 | 2010 | The Kuznets-Kaldor-Puzzle and Neutral Cross-Capital-Intensity Structural Change | Stijepic, Denis Wagner, Helmut |
| 453 | 2010 | Monetary Policy and Boom-Bust Cycles: The Role of Communication | Knütter, Rolf Wagner, Helmut |
| 454 | 2010 | Konsensuale Effizienzbewertung und –verbesserung mittels DEA – Output- vs. Inputorientierung – | Reucher, Elmar Röder, Wilhelm |
| 455 | 2010 | Consistent Modeling of Risk Averse Behavior with Spectral Risk Measures | Wächter, Hans Peter Mazzoni, Thomas |

| | | | |
|-----|------|--|---|
| 456 | 2010 | Der virtuelle Peer – Eine Anwendung der DEA zur konsensualen Effizienz- bewertung – | Reucher, Elmar |
| 457 | 2010 | A two-stage packing procedure for a Portuguese trading company | Moura, Ana Bortfeldt, Andreas |
| 458 | 2010 | A tree search algorithm for solving the multi-dimensional strip packing problem with guillotine cutting constraint | Bortfeldt, Andreas Jungmann, Sabine |
| 459 | 2010 | Equity and Efficiency in Regional Public Good Supply with Imperfect Labour Mobility – Horizontal versus Vertical Equalization | Arnold, Volker |
| 460 | 2010 | A hybrid algorithm for the capacitated vehicle routing problem with three-dimensional loading constraints | Bortfeldt, Andreas |
| 461 | 2010 | A tree search procedure for the container relocation problem | Forster, Florian Bortfeldt, Andreas |
| 462 | 2011 | Advanced X-Efficiencies for CCR- and BCC-Modell – Towards Peer-based DEA Controlling | Rödder, Wilhelm Reucher, Elmar |
| 463 | 2011 | The Effects of Central Bank Communication on Financial Stability: A Systematization of the Empirical Evidence | Knütter, Rolf Mohr, Benjamin Wagner, Helmut |
| 464 | 2011 | Lösungskonzepte zur Allokation von Kooperationsvorteilen in der kooperativen Transportdisposition | Strangmeier, Reinhard Fiedler, Matthias |
| 465 | 2011 | Grenzen einer Legitimation staatlicher Maßnahmen gegenüber Kreditinstituten zur Verhinderung von Banken- und Wirtschaftskrisen | Merbecks, Ute |
| 466 | 2011 | Controlling im Stadtmarketing – Eine Analyse des Hagener Schaufensterwettbewerbs 2010 | Fließ, Sabine Bauer, Katharina |
| 467 | 2011 | A Structural Approach to Financial Stability: On the Beneficial Role of Regulatory Governance | Mohr, Benjamin Wagner, Helmut |
| 468 | 2011 | Data Envelopment Analysis - Skalenerträge und Kreuzskalenerträge | Wilhelm Rödder Andreas Dellnitz |
| 469 | 2011 | Controlling organisatorischer Entscheidungen: Konzeptionelle Überlegungen | Lindner, Florian Scherer, Ewald |
| 470 | 2011 | Orientierung in Dienstleistungsumgebungen – eine explorative Studie am Beispiel des Flughafen Frankfurt am Main | Fließ, Sabine Colaci, Antje Nesper, Jens |

| | | | |
|-----|------|---|---|
| 471 | 2011 | Inequality aversion, income skewness and the theory of the welfare state | Weinreich, Daniel |
| 472 | 2011 | A tree search procedure for the container retrieval problem | Forster, Florian Bortfeldt, Andreas |
| 473 | 2011 | A Functional Approach to Pricing Complex Barrier Options | Mazzoni, Thomas |
| 474 | 2011 | Bologna-Prozess und neues Steuerungsmodell – auf Konfrontationskurs mit universitären Identitäten | Jost, Tobias Scherer, Ewald |
| 475 | 2011 | A reduction approach for solving the rectangle packing area minimization problem | Bortfeldt, Andreas |
| 476 | 2011 | Trade and Unemployment with Heterogeneous Firms: How Good Jobs Are Lost | Altenburg, Lutz |
| 477 | 2012 | Structural Change Patterns and Development: China in Comparison | Wagner, Helmut |
| 478 | 2012 | Demografische Risiken – Herausforderungen für das finanzwirtschaftliche Risikomanagement im Rahmen der betrieblichen Altersversorgung | Merbecks, Ute |
| 479 | 2012 | “It’s all in the Mix!” – Internalizing Externalities with R&D Subsidies and Environmental Liability | Endres, Alfred Friehe, Tim Rundshagen, Bianca |
| 480 | 2012 | Ökonomische Interpretationen der SkalenvARIABLEN u in der DEA | Dellnitz, Andreas Kleine, Andreas Rödter, Wilhelm |
| 481 | 2012 | Entropiebasierte Analyse von Interaktionen in Sozialen Netzwerken | Rödter, Wilhelm Brenner, Dominic Kulmann, Friedhelm |
| 482 | 2013 | Central Bank Independence and Financial Stability: A Tale of Perfect Harmony? | Berger, Wolfram Kißner, Friedrich |
| 483 | 2013 | Energy generation with Directed Technical Change | Kollenbach, Gilbert |
| 484 | 2013 | Monetary Policy and Asset Prices: When Cleaning Up Hits the Zero Lower Bound | Berger, Wolfram Kißner, Friedrich |
| 485 | 2013 | Superknoten in Sozialen Netzwerken – eine entropieoptimale Analyse | Brenner, Dominic, Rödter, Wilhelm, Kulmann, Friedhelm |