

Latent-Space Network Models

23.07.2017



Autor: Florian Pargent

Inhaltsverzeichnis

1	Grundidee des Latent-Space Ansatzes	3
1.1	Typische Eigenschaften von Netzwerken	3
1.2	Modellierung eines latenten Social-Space	3
2	Beispieldatensatz – Publikationen am Institut für Statistik	4
3	Latent-Space Modelle und deren Anwendung	5
3.1	Latent Position Model	6
3.2	Latent Cluster Model	8
3.3	Latent Cluster Random Effects Model	9
4	Erweiterung auf Zähldaten	13
5	Schätzung der Modellparameter	15
5.1	Maximum-Likelihood Schätzung am Beispiel des LPM	15
5.2	Volle Bayes-Inferenz am Beispiel des LCM	16
5.2.1	Priori Verteilungen	16
5.2.2	Vollständig Bedingte Verteilungen	16
5.2.3	Metropolis-within Gibbs MCMC-Algorithmus	17
5.3	Identifizierbarkeit der Modellparameter	18
5.3.1	Latente Positionen	18
5.3.2	Clusterlabels	19
5.3.3	Intercept/Random Effects im LCREM	19
	Quellen	19

1 Grundidee des Latent-Space Ansatzes

1.1 Typische Eigenschaften von Netzwerken

Soziale Netzwerke weisen eine Reihe charakteristischer Eigenschaften auf (Krivitsky et al. 2009):

- *Transitivity*: Eine Kante zwischen zwei Knoten ist wahrscheinlicher, wenn beide Knoten einen gemeinsamen Nachbarn teilen.
- *Homophily on Observed Attributes*: Eine Kante zwischen zwei Knoten ist wahrscheinlicher, wenn sich beide Knoten hinsichtlich bestimmter Charakteristiken ähnlich sind.
- *Community Structure*: Knoten im Netzwerk gruppieren sich in Clustern, innerhalb derer Kanten wahrscheinlicher sind als zwischen verschiedenen Clustern.
- *Degree Heterogeneity*: Manche Knoten im Netzwerk weisen besonders viele, andere besonders wenige Kanten auf.

Klassische statistische Modelle wie generalisierte lineare Modelle (GLMs) sind nicht in der Lage, diese Eigenschaften von Netzwerken sinnvoll darzustellen. Im Folgenden wird eine Modellklasse dargestellt, die für die statistische Analyse von Netzwerken besser geeignet ist.

1.2 Modellierung eines latenten Social-Space

Die Grundidee sogenannter *Latent-Space Modelle* für Netzwerkdaten (Hoff, Raftery, and Handcock 2002) besteht in der Modellierung eines latenten *Social-Space*. Dabei wird jeder Knoten im Netzwerk durch einen Positionsvektor im \mathbb{R}^d repräsentiert. Typischerweise wählt man $d = 2$, in seltenen Fällen auch $d \geq 3$. *Latent-Space Modelle* modellieren die Wahrscheinlichkeit für das Vorliegen einer Kante zwischen zwei Knoten im Netzwerk, bedingt auf die latenten Positionen der beteiligten Knoten. Dabei wird angenommen, dass diese Wahrscheinlichkeit vom euklidischen Abstand der beiden Knoten im *Social-Space* abhängt und alle Kanten, gegeben der latenten Positionen der Knoten, voneinander unabhängig sind.

Wie wir später anhand der einzelnen *Latent-Space Modelle* sehen werden, weisen diese durch geeignete Erweiterungen der *Social-Space* Idee alle der beschriebenen Eigenschaften sozialer Netzwerke auf. Außerdem liefern diese Modelle, vor allem bei Verwendung eines zweidimensionalen *Social-Space*, eine natürliche grafische Darstellung des Netzwerks. Diese Repräsentation kann vom Anwender zur explorativen Analyse und Interpretation genutzt werden.

2 Beispieldatensatz – Publikationen am Institut für Statistik

Vor der Darstellung konkreter *Latent-Space Modelle*, wird zunächst ein realer Netzwerkdatensatz vorgestellt, anhand dessen alle Modelle exemplarisch erläutert werden:



Abbildung 1: Screenshot der Homepage des Instituts für Statistik

Das Institut für Statistik der Ludwig-Maximilians-Universität München unterhält auf seiner Homepage eine Liste wissenschaftlicher Publikationen der wissenschaftlichen Mitarbeiter am Institut (siehe Abbildung 1). Diese Liste umfasst sowohl Publikationen in wissenschaftlichen Fachzeitschriften, als auch über die Universitätsbibliothek veröffentlichte Working-Papers. Dabei müssen Mitarbeiter ihre Publikationen explizit melden, damit diese in die Liste aufgenommen werden. Daher liegen bei weitem nicht alle Publikationen des Instituts vor. Außerdem scheint es, als würden manche Arbeitsgruppen einen höheren Anteil ihrer Publikationen melden als andere.

Mithilfe des R-Pakets `rvest` wurden am 31.05.2017 alle auf der Homepage geführten Publikationen mitsamt ihren Autoren automatisch ausgelesen und mit der ebenfalls online verfügbaren Liste aller Mitarbeiter des Institut abgeglichen. Das Ergebnis ist ein Netzwerkdatensatz, der für alle Paare von Mitarbeitern angibt, wie viele gemeinsame Publikationen diese jeweils aufweisen. Die Kanten im Netzwerk werden für alle Analysen

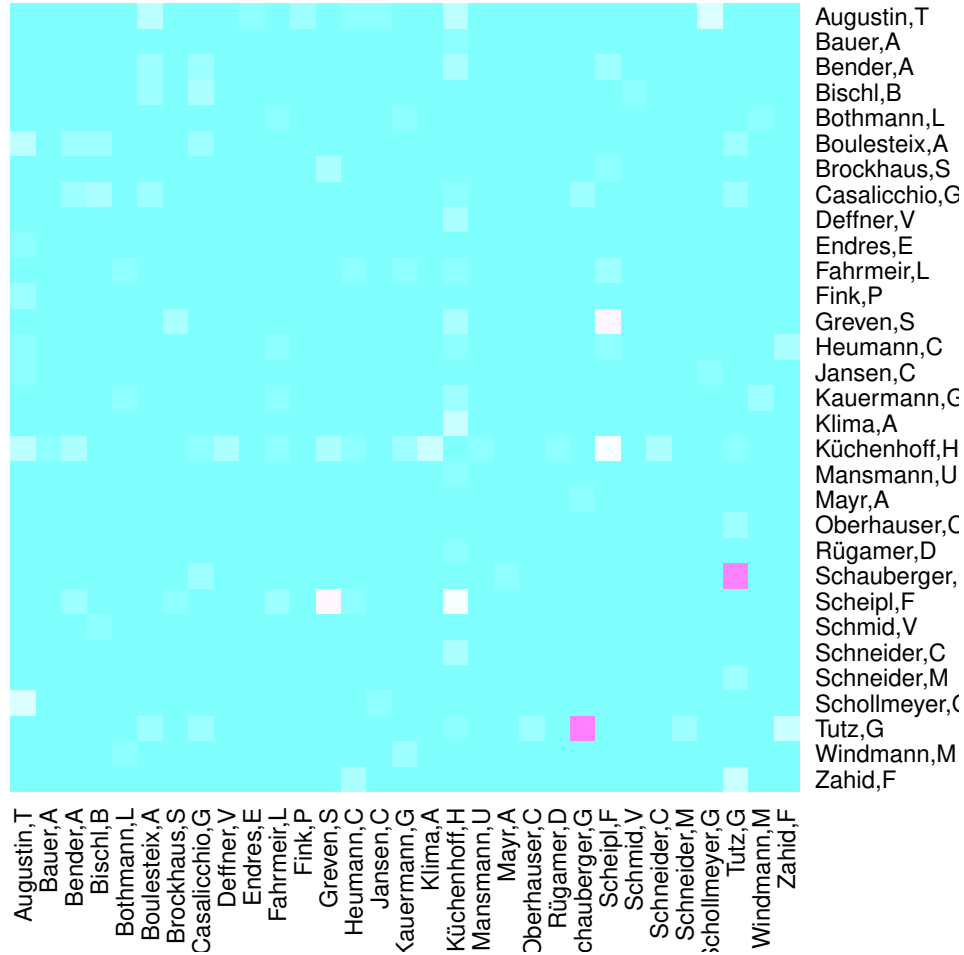


Abbildung 2: Heatmap des Publikationsdatensatzes

als ungerichtet angenommen.

Abbildung 2 veranschaulicht die Soziomatrix des Publikationsdatensatzes in Form einer Heatmap. Der Datensatz enthält 31 Personen. Wissenschaftliche Mitarbeiter des Instituts ohne gemeinsame Publikationen mit anderen Mitarbeitern wurden zuvor entfernt. Es fällt auf, dass Prof. Helmut Küchenhoff besonders viele Publikationen mit einer großen Zahl verschiedener Publikationspartner aufweist. Einen weiteren Ausreißerwert stellen die 17 gemeinsamen Publikationen zwischen Prof. Gerhard Tutz und Gunther Schauburger dar.

3 Latent-Space Modelle und deren Anwendung

Im Folgenden werden verschiedene *Latent-Space Modelle* in aufsteigender Komplexität vorgestellt. Jedes Modell wird anhand des Publikationsdatensatzes veranschaulicht. In den meisten Modellen wird nur berücksichtigt, ob eine bestimmte Kante zwischen zwei Knoten vorliegt oder nicht. In diesen Fällen wird der Publikationsdatensatz

in ein binäres Netzwerk umgewandelt. Die Information, wie viele Publikationen ein Mitarbeiterpaar aufweist geht dementsprechend verloren. Zur Schätzung der Modelle wurde jeweils das R-Paket `latentnet` verwendet (Krivitsky and Handcock 2008). Nähere Informationen zur Modellschätzung folgen in Kapitel 5.

3.1 Latent Position Model

Das einfachste Latent-Space Modell ist das von Hoff, Raftery, and Handcock (2002) entwickelte *Latent-Position Model* (LPM):

$$\log \left(\frac{P(Y_{ij} = 1|Z, x, \beta)}{P(Y_{ij} = 0|Z, x, \beta)} \right) = \sum_{k=0}^p \beta_k x_{kij} - \|Z_i - Z_j\|$$

Im LPM wird die Wahrscheinlichkeit für das Vorliegen einer Kante im Netzwerk zwischen den Knoten i und j modelliert ($Y_{ij} \in \{0, 1\}$), bedingt auf die latenten Positionen Z . Diese Wahrscheinlichkeit kann potentiell von einer Reihe von p Kovariablen x_{kij} abhängen, die jeweils die Kante zwischen den Knoten i und j charakterisieren.

Das LPM weist sehr starke Ähnlichkeit mit der logistischen Regression auf. Die Modellgleichungen unterscheiden sich nur dadurch, dass der lineare Prädiktor zusätzlich zum Intercept β_0 und den Effekten der Kovariablen x_{kij} den euklidischen Abstand zwischen den d -dimensionalen Positionsvektoren Z_i und Z_j der beteiligten Knoten enthält. Durch die Einführung der Positionsvektoren ist das LPM in der Lage *Transitivity* im Netzwerk abzubilden. Knoten mit gemeinsamen Nachbarn liegen im *Social-Space* näher beieinander und weisen durch den negativen Effekt des euklidischen Abstands eine höhere Wahrscheinlichkeit auf, durch eine Kante verbunden zu sein. Das LPM bildet auch *Homophily on Observed Attributes* ab. Einerseits können Ähnlichkeiten zwischen Knoten durch beobachtbare Kovariablen in das Modell aufgenommen werden. Andererseits können nicht beobachtete Ähnlichkeiten ebenfalls durch eine räumliche Gruppierung von Knoten mit hoher Interkonnektivität berücksichtigt werden.

Abbildung 3 zeigt die geschätzten Positionsvektoren des LPM für den Publikationsdatensatz. Beobachtete Kanten im Datensatz werden durch graue Linien symbolisiert. Für dieses, sowie alle folgenden Modelle wurde ein zweidimensionaler *Social-Space* modelliert. Die Achsen mit den beiden Koordinaten der Positionsvektoren wurden aus Gründen der Übersichtlichkeit ausgeblendet.

LPM

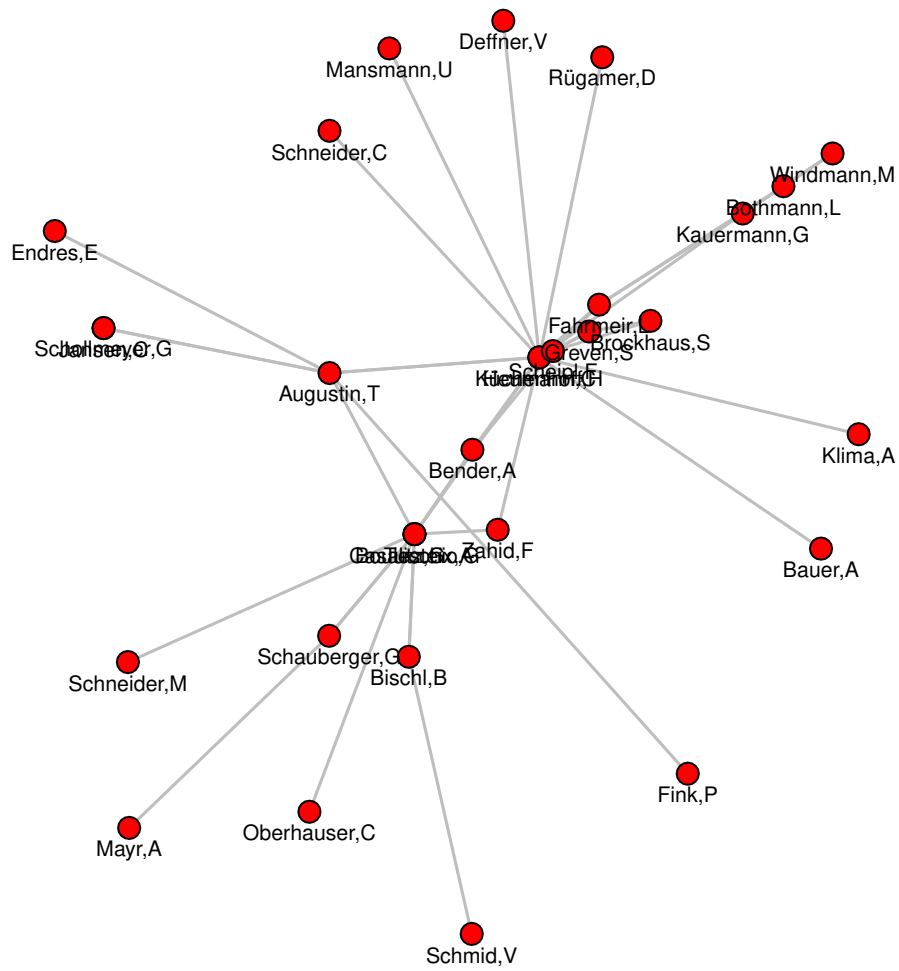


Abbildung 3: LPM am Beispiel des Publikationsdatensatzes

3.2 Latent Cluster Model

Eine Erweiterung des LPM stellt das *Latent-Cluster Model* (LCM) dar (Handcock, Raftery, and Tantrum 2007):

$$\log \left(\frac{P(Y_{ij} = 1|Z, x, \beta)}{P(Y_{ij} = 0|Z, x, \beta)} \right) = \sum_{k=0}^p \beta_k x_{kij} - \|Z_i - Z_j\|$$

$$Z_i \sim \sum_{g=1}^G \lambda_g MVN_d(\mu_g, \sigma_g^2 I_d)$$

Im Unterschied zum LPM wird im LCM angenommen, dass die Positionsvektoren aus einer Mischung von G Normalverteilungen mit Dimension d stammen, deren Mischungsverhältnis durch den Parameter λ_g gesteuert wird. Typischerweise werden die einzelnen Komponenten der multivariaten Normalverteilungen als unkorreliert angenommen. Eine Gruppe von Knoten, die untereinander viele Kanten aufweisen haben im LCM die Möglichkeit ein Cluster zu bilden. Damit ist das LCM in der Lage typische Formen von *Community Structure* abzubilden.

In Abbildung 4 wird das geschätzte LCM mit drei Clustern für den Publikationsdatensatz visualisiert. Jedes Cluster wird in einer eigenen Farbe dargestellt, mit dem geschätzten Erwartungswert des Clusters $\hat{\mu}_g$ als Kreuz, umgeben von einem Kreis mit Radius $\hat{\sigma}_g$. Für jeden Knoten im Netzwerk werden mithilfe eines eingefärbten Kuchendiagramms die geschätzten Clusterwahrscheinlichkeiten abgebildet. Auffällig ist, dass im Publikationsdatensatz mithilfe des LCM keine Clusterstruktur resultiert. Vergleicht man Modelle mit unterschiedlich vielen Clustern mithilfe einer Variante des BIC (Handcock, Raftery, and Tantrum 2007), so scheint ein LCM mit nur einem Cluster die Daten am besten zu beschreiben. Dies ist auf den ersten Blick verwunderlich, da sich deskriptiv durchaus eine stärkere Vernetzung innerhalb einzelner Arbeitsgruppen des Instituts erkennen lässt.

LCM

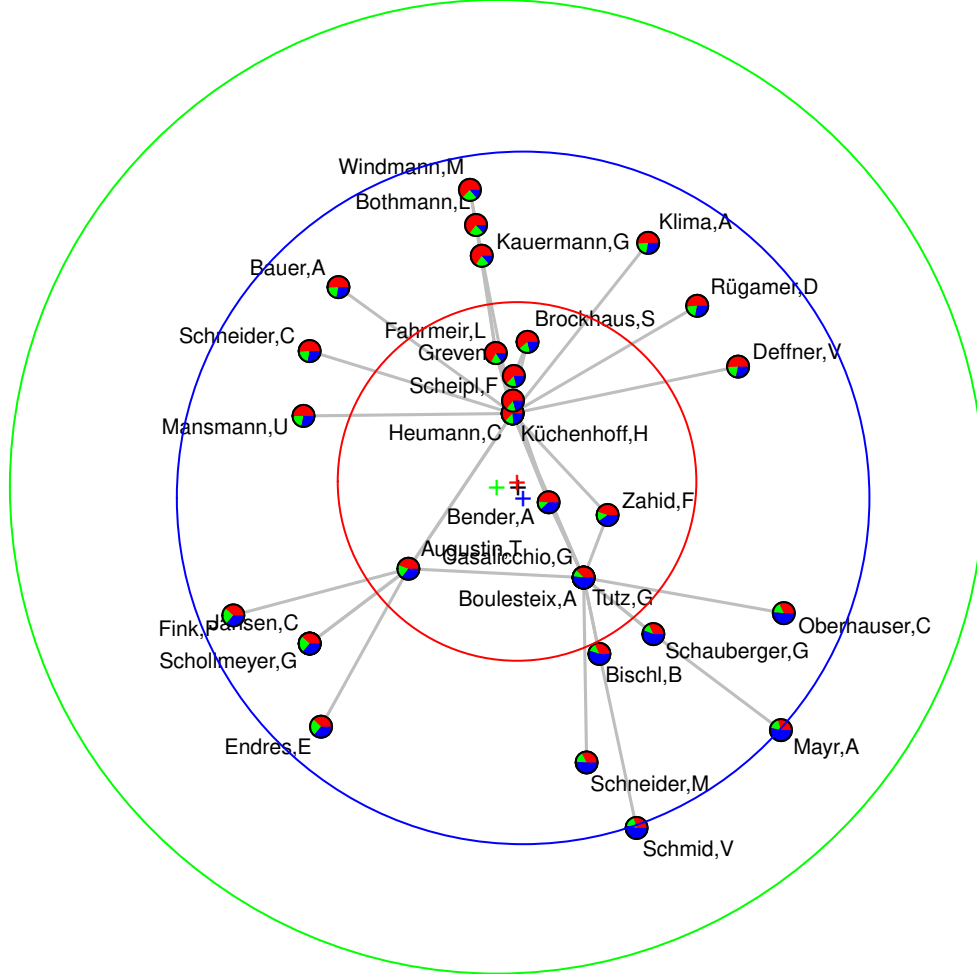


Abbildung 4: LCM mit drei Clustern am Beispiel des Publikationsdatensatzes

3.3 Latent Cluster Random Effects Model

Krivitsky et al. (2009) stellen mit dem *Latent-Cluster Random Effects Model* (LCREM) eine Erweiterung des LCM vor:

$$\log \left(\frac{P(Y_{ij} = 1 | Z, x, \beta)}{P(Y_{ij} = 0 | Z, x, \beta)} \right) = \sum_{k=0}^p \beta_k x_{kij} - \|Z_i - Z_j\| + \delta_i + \delta_j$$

$$\delta_i \sim N(0, \sigma_\delta^2)$$

$$Z_i \sim \sum_{g=1}^G \lambda_g MVN_d(\mu_g, \sigma_g^2 I_d)$$

Der lineare Prädiktor des LCREM enthält zusätzlich zu den bisherigen Modellen Random Effects (δ_i und δ_j) für beide an der jeweiligen Kante beteiligten Knoten. Alle Random Effects werden als *iid*-normalverteilt angenommen. Durch die Einführung der Random Effects ist es möglich, Knoten unabhängig von deren Position im *Social-Space* eine höhere oder niedrigere Tendenz zuzuschreiben, mit anderen Knoten Verbindungen einzugehen. Damit ist das LCREM in der Lage *Degree Heterogeneity* zu modellieren.

Die Auswirkung dieser Modifikation für die Modellierung des Publikationsdatensatzes wird in Abbildung 5 ersichtlich, welche ein geschätztes LCREM mit der BIC-optimalen Zahl von vier Clustern zeigt. Die Cluster spiegeln größtenteils einzelne Arbeitsgruppen des Instituts wider. Diese Struktur scheint sich dadurch zu ergeben, dass Doktoranden und andere wissenschaftliche Mitarbeiter in der Regel nur mit dem Leiter ihrer Arbeitsgruppe gemeinsam publizieren, während die Professoren des Instituts häufig auch untereinander Publikationen aufweisen. Im Rahmen des LCM scheint diese Clusterbildung nicht möglich, da die Professoren ohne die Einführung der Random Effects zusammen gruppiert werden müssen, um deren große Tendenz für gemeinsame Publikationen abzubilden.

Alle dargestellten *Latent-Space Modelle* sind in der Lage, externe Kovariablen mit einzubeziehen. Betrachtet man Abbildung 5 genauer, so fällt auf, dass die auf der Homepage des Instituts für Statistik geführten Professoren (Augustin,T; Bischl,B; Boulesteix,A; Fahrmeir,L; Greven,S; Heumann,C; Kauermann,G; Küchenhoff,H; Mansmann,U; Mayr,A; Schmid,V; Tutz,G) tendentiell höhere Random Effects aufweisen, da Sie im Vergleich zu den anderen Mitarbeitern besonders viel publizieren. Um diese zusätzliche Information zu berücksichtigen, wurde die dreistufige Kovariable *Professoren* erstellt. Diese gibt für jede Kante im Netzwerk an, ob an der Kante *ein*, *zwei*, oder *kein* Knoten mit Professorentitel beteiligt sind.

Abbildung 6 zeigt erneut ein LCREM mit vier Clustern, in das die Kovariable *Professoren* in Form zweier Dummyvariablen aufgenommen wurde. Während die Anordnung der Knoten im *Social-Space* nahezu unverändert bleibt, fallen die Unterschiede in den Random Effects nun deutlich schwächer aus. Tabelle 1 zeigt die Parameterschätzungen für die Kovariable *Professoren*, wobei die Ausprägung *Ein Professor* als Referenzkategorie gewählt wurde.

LCREM

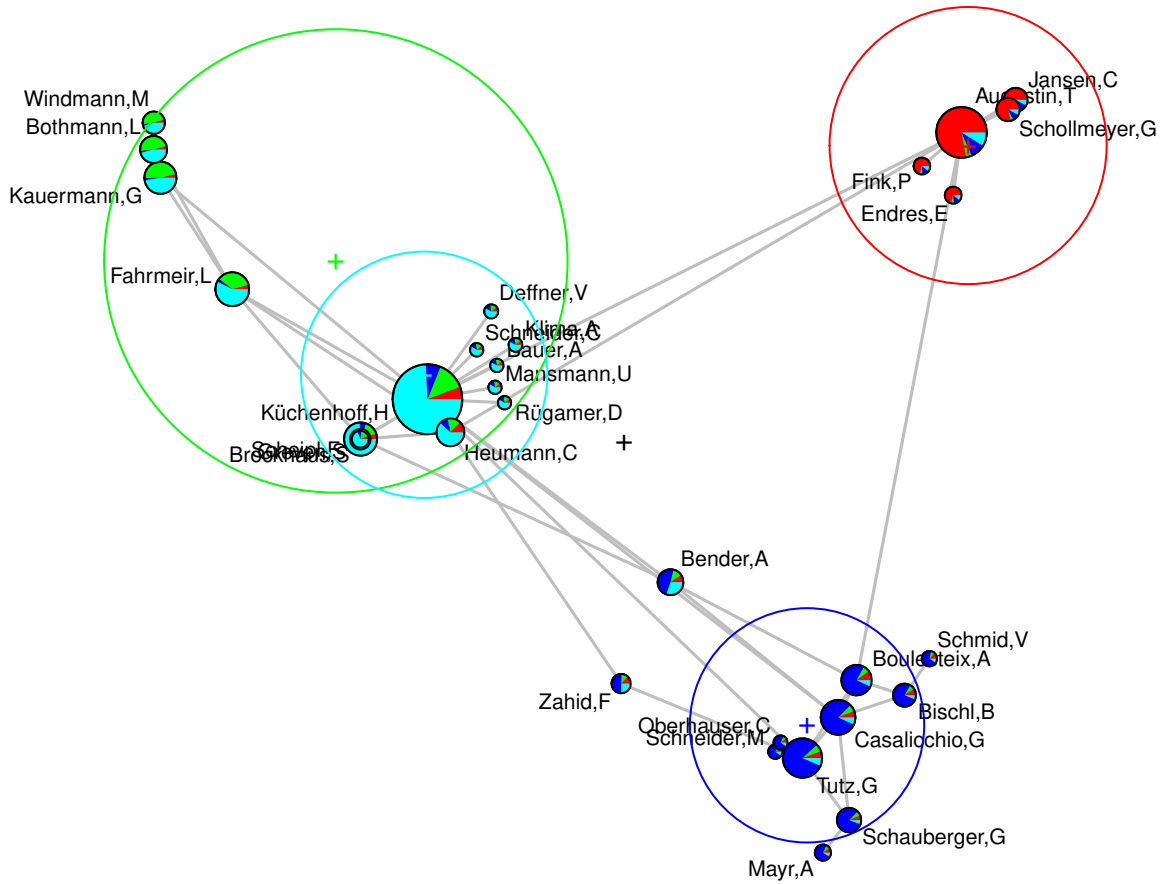


Abbildung 5: LCREM mit vier Clustern am Beispiel des Publikationsdatensatzes

Tabelle 1: Parameterschätzungen im LCREM des Publikationsdatensatzes für die Kovariable Professoren

	Posteriori Mittelwert	2.5%	97.5%
Intercept (Ein Professor)	0.37	-0.32	1.12
Kein Professor	-1.81	-3.42	-0.32
Zwei Professoren	1.40	-0.06	2.93

LCREM mit Kovariable Professoren

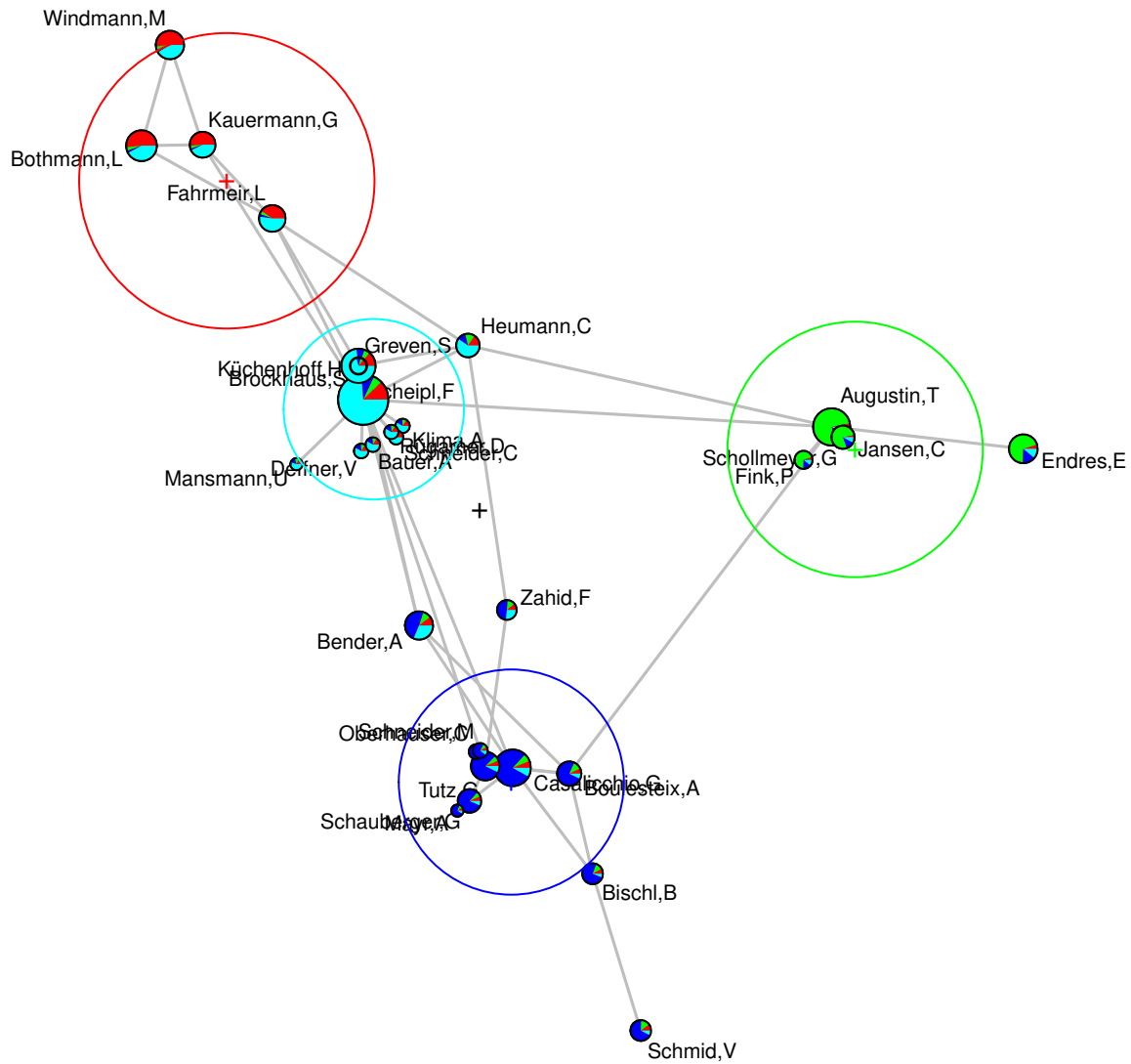


Abbildung 6: LCREM mit vier Clustern und der Kovariable Professoren am Beispiel des Publikationsdatensatzes

4 Erweiterung auf Zähldaten

Alle *Latent-Space Modelle* lassen sich sehr leicht auf nicht binäre Netzwerkdaten erweitern. Ein häufiger Anwendungsfall sind Netzwerke mit gewichteten Kanten in Form von Zählvariablen, die in der Regel mit einer Poissonverteilung modelliert werden. Für das entsprechende LCREM ergibt sich analog zu den bekannten GLMs die folgende Form:

$$\begin{aligned}
 Y_{ij} | \mu_{ij} &\sim \text{Poisson}(\mu_{ij}) \\
 \log(\mu_{ij}) &= \sum_{k=0}^p \beta_k x_{kij} - \|Z_i - Z_j\| + \delta_i + \delta_j \\
 \delta_i &\sim N(0, \sigma_\delta^2) \\
 Z_i &\sim \sum_{g=1}^G \lambda_g \text{MVN}_d(\mu_g, \sigma_g^2 I_d)
 \end{aligned}$$

Für den bedingten Erwartungswert wird als Verteilungsform eine Poissonverteilung angenommen. Wie bei den GLMs mit Poisson-Response üblich, wird hier der Logarithmus als Linkfunktion verwendet, während bei den Modellen für binäre Netzwerke die logistische Linkfunktion zur Anwendung kommt.

In Abbildung 7 wurde für den Publikationsdatensatz ein LCREM mit Poisson Response geschätzt, wobei drei Cluster spezifiziert und die Kovariable *Professoren* erneut mit in das Modell aufgenommen wurde. Die Parameterschätzungen finden sich in Tabelle 2. Im Gegensatz zu den bisherigen Auswertungen wird in diesem Modell nicht nur berücksichtigt, ob zwei Mitarbeiter bereits gemeinsam publiziert haben, sondern auch wie viele gemeinsame Publikationen für das jeweilige Paar vorliegen. Wie aufgrund der extrem hohen Anzahl von gemeinsamen Publikationen zu erwarten, sind die Knoten von Prof. Gerhard Tutz und Gunther Schauburger direkt nebeneinander angeordnet und erhalten zusätzlich noch hohe Random Effects. Allerdings erwies sich das Modell für Zähldaten insgesamt als sehr instabil und sollte daher nur mit großer Vorsicht interpretiert werden.

LCREM mit Poisson Response und Kovariable Professoren

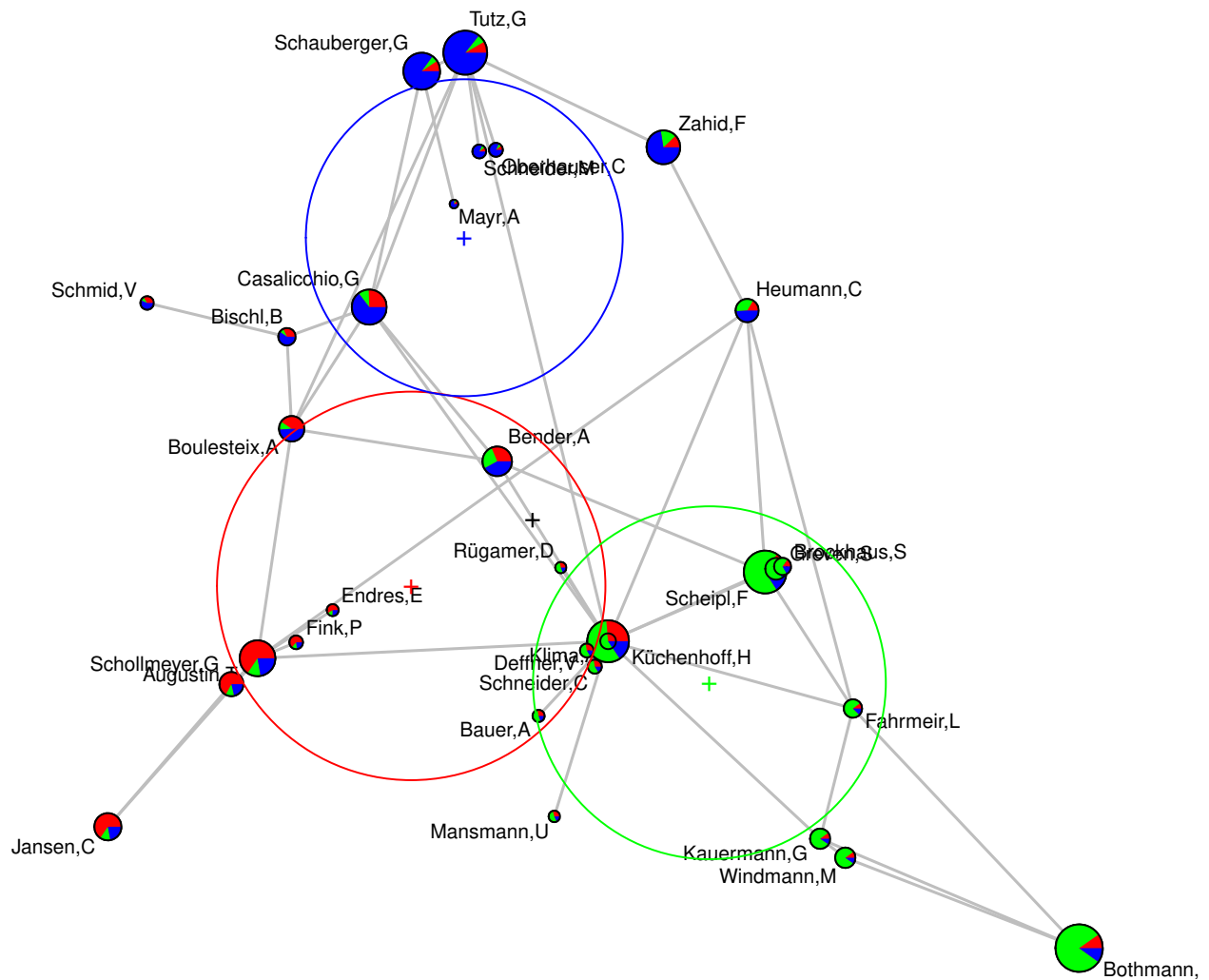


Abbildung 7: LCREM mit vier Clustern und der Kovariable Professoren am Beispiel des Publikationsdatensatzes. Modellierung der Anzahl der Publikationen als Poissonvariable

Tabelle 2: Parameterschätzungen im LCREM mit Poisson Response des Publikationsdatensatzes für die Kovariable Professoren

	Posteriori Mittelwert	2.5%	97.5%
Intercept (Ein Professor)	0.37	-0.32	1.12
Kein Professor	-1.81	-3.42	-0.32
Zwei Professoren	1.40	-0.06	2.93

5 Schätzung der Modellparameter

5.1 Maximum-Likelihood Schätzung am Beispiel des LPM

Die Likelihood (und log-Likelihood) des LPM für binäre Netzwerke hat die gleiche Struktur wie die Likelihood der logistischen Regression. Beide unterscheiden sich nur in Form des linearen Prädiktors η .

$$\log(f(y|\eta)) = \sum_{i \neq j} (\eta_{ij} y_{ij} - \log(1 + e^{\eta_{ij}}))$$

$$\eta_{ij} = \sum_{k=0}^p \beta_k x_{kij} - \|Z_i - Z_j\| = \sum_{k=0}^p \beta_k x_{kij} - d_{ij}$$

Betrachtet man den linearen Prädiktor genauer, so stellt man fest, dass die Likelihood des LPM konkav ist bezüglich der Distanzen d_{ij} , jedoch nicht hinsichtlich der latenten Positionen Z_i . Damit ist das bei der Maximum-Likelihood Schätzung resultierende Optimierungsproblem komplizierter als bei der logistischen Regression. Es besteht keine Garantie, dass ein globales Maximum numerisch gefunden werden kann, oder ein solches überhaupt existiert.

Zur Optimierung der Likelihood muss daher auf universelle Algorithmen für nicht-lineare Optimierung zurück gegriffen werden. Dabei sollten multiple Startwerte sinnvoll gewählt werden. Ein praktisch realisierbares Vorgehen sieht dabei folgendermaßen aus (Handcock, Raftery, and Tantrum 2007):

- Approximiere die Distanzmatrix D durch geodesische Distanzen
- Identifiziere die Positionsmatrix Z aus der Distanzmatrix D mit multidimensionaler Skalierung
- Verwende die Positionsmatrix Z als Startwerte für nicht-lineare Optimierungsalgorithmen, wie sie beispielsweise in der Funktion `optim` in R implementiert sind

5.2 Volle Bayes-Inferenz am Beispiel des LCM

Aufgrund der dargestellten Probleme bei der Likelihood-Inferenz werden *Latent-Space Modelle* in den gängigen R-Paketen üblicherweise mit voller Bayes-Inferenz geschätzt. So bietet sich etwa für das LCM ein klassischer Metropolis-within-Gibbs MCMC-Algorithmus an, der im Folgenden kurz vorgestellt wird:

5.2.1 Priori Verteilungen

Zunächst werden für die Modellparameter $\beta = (\beta_0, \beta_1, \dots, \beta_p)$, $\lambda = (\lambda_1, \dots, \lambda_G)$, σ_g^2 und μ_g^2 Priori Verteilungen spezifiziert.

$$\beta \sim MVN_{p+1}(\xi, \Psi)$$

$$\lambda \sim Dirichlet(\nu)$$

$$\mu_g \sim MVN_d(0, \omega^2 I_d)$$

$$\sigma_g^2 \sim \sigma_0^2 Inv\chi_\alpha^2$$

Handcock, Raftery, and Tantrum (2007) wählen für die Hyperparameter $\xi = 0$, $\Psi = 2I$, $\nu_g = 3$, $\omega^2 = 2$, $\sigma_0^2 = 0.103$ und $\alpha = 2$.

5.2.2 Vollständig Bedingte Verteilungen

Betrachtet man die vollständig bedingten Verteilungen der Modellparameter, so stellt man fest, dass diese für Z_i sowie β bis auf Konstanten bekannt sind.

$$z_i | K_i, \dots \propto \phi_d(z_i; \mu_g, \sigma_g^2 I_d) f(y | Z, x, \beta)$$

$$\beta | Z, \dots \propto \phi_{p+1}(\beta; \xi, \Psi) f(y | Z, x, \beta)$$

Dabei steht ϕ_d für die Dichtefunktion der d -dimensionalen Normalverteilung.

Für die restlichen Parameter sind die vollständig bedingten Verteilungen aufgrund der semi-konjugierten Priori Verteilungen sogar vollständig bekannt.

$$\lambda | \dots \sim Dirichlet(m + \nu)$$

$$\mu_g | \dots \sim MVN_d \left(\frac{m_g \bar{z}_g}{m_g + \sigma_g / \omega^2}, \frac{\sigma_g^2}{m_g + \sigma_g^2 / \omega^2} I \right)$$

$$\sigma_g^2 | \dots \sim (\sigma_0^2 + d s_g^2) Inv \chi_{\alpha + m_g d}^2$$

$$P(K_i = g | \dots) = \frac{\lambda_g \phi_d(z_i; \mu_g, \sigma_g^2 I_d)}{\sum_{r=1}^G \lambda_r \phi_d(z_i; \mu_r, \sigma_r^2 I_d)}$$

5.2.3 Metropolis-within Gibbs MCMC-Algorithmus

Es ergibt sich der folgende Metropolis-within Gibbs MCMC-Algorithmus:

Metropolis-Algorithmus für Z_{t+1}

- Vorschlag: $Z_i^* \sim MVN_d(Z_{it}, \tau_Z^2 I_d)$
- Akzeptanz WK: $\frac{f(y|Z_{t+1}, x, \beta_t) \phi_d(Z_i^*; \mu_{K_i}, \sigma_{K_i}^2 I_d)}{f(y|Z_t, x, \beta_t) \phi_d(Z_{it}; \mu_{K_i}, \sigma_{K_i}^2 I_d)}$

Dabei werden die Positionsvektoren Z_i in zufälliger Reihenfolge upgedatet.

Metropolis-Algorithmus für β_{t+1}

- Vorschlag: $\beta^* \sim MVN_{p+1}(\beta_t, \tau_\beta^2 I_{p+1})$
- Akzeptanz WK: $\frac{f(y|Z_{t+1}, x, \beta^*) \phi_{p+1}(\beta^*; \xi, \Psi)}{f(y|Z_{t+1}, x, \beta_t) \phi_{p+1}(\beta_t; \xi, \Psi)}$

Sowohl für Z_i als auch β werden symmetrische Random-Walk Vorschlagsdichten verwendet. Daher reduziert sich der Metropolis-Hastings Algorithmus zum Metropolis Algorithmus. In diesem wird jeder Vorschlag mit einer, bis auf Konstanten, größeren Dichte als der aktuelle Parameterwert sicher angenommen.

Gibbs-Sampling für K_i, μ_g, σ_g^2 und λ_g

Da für diese Parameter die vollständig bedingten Dichten vollständig bekannt sind, kann direkt aus ihnen gezogen werden.

Bei dem hier dargestellten Metropolis-within Gibbs MCMC-Algorithmus sind die Ziehungen der latenten Positionen in der Regel sehr hoch mit den Clusterzugehörigkeiten korreliert, weshalb für eine stabile Schätzung sehr viele Iterationen notwendig sind (Handcock, Raftery, and Tantrum 2007). Durch die Einführung der Random Effects im LCREM ist außerdem davon auszugehen, dass die Effektivität des Metropolis-within Gibbs Algorithmus weiter abnimmt. Daher wird für die Modellschätzung im R-Paket `latentnet` ein modifizierter

MCMC-Algorithmus verwendet, der mit korrelierten Vorschlagsdichten, verschiedenen Block-Updates und automatisch kalibrierten Startwerten arbeitet. Die technischen Details finden sich in Krivitsky et al. (2009). Für die Schätzung der *Latent-Space Modelle* anhand des Publikationsdatensatzes wurden nach einem Burnin von 60000, 30000 effektive Ziehungen im Abstand von jeweils 30 Iterationen gezogen. Als Hyperparameter der Priori Verteilungen wurden die Default-Einstellungen verwendet (Krivitsky and Handcock 2008).

5.3 Identifizierbarkeit der Modellparameter

In *Latent-Space Modellen* bestehen einige Besonderheiten hinsichtlich der Identifizierbarkeit der Modellparameter, die bei der Inferenz berücksichtigt werden sollten:

5.3.1 Latente Positionen

Die Likelihood aller *Latent-Space Modelle* hängt von den latenten Positionen Z_i nur über deren Distanzen ab. Damit ist die Likelihood invariant gegenüber Spiegelung, Rotation und Verschiebung der latenten Positionen.

Optimale Punktschätzer in der Bayes-Inferenz minimieren den a-posteriori erwarteten Verlust bezüglich einer gewählten Verlustfunktion. Wählt man beispielsweise die quadratische Verlustfunktion $L_2(\eta, \eta^*) = (\eta - \eta^*)^2$, so ergibt sich als optimaler Bayes-Schätzer der Posteriori Erwartungswert. Shortreed, Handcock, and Hoff (2006) zeigen, dass der Posteriori Erwartungswert wie auch der Posteriori Modus unter den dargestellten Invarianzeigenschaften nicht gut geeignet sind, um die oftmals multi-modale Posterioriverteilung sinnvoll zu beschreiben. Sie schlagen daher einen anderen Punktschätzer $\hat{\eta}_{MKL}$ vor, bei dem der a-posteriori erwartete Verlust bezüglich der Kullback-Leibler Verlustfunktion $KL(\eta, \eta_*) = E_\eta \left(\log \left(\frac{f(y|\eta)}{f(y|\eta_*)} \right) \right)$ minimiert wird. Es resultiert das folgende Optimierungsproblem:

$$\begin{aligned} \hat{\eta}_{MKL} &= \underset{\eta_*}{\operatorname{argmin}} E_{\eta|Y_{obs}}(KL(\eta, \eta_*)) \\ &= \underset{\eta_*}{\operatorname{argmax}} \frac{\exp(\eta_*^T E_\eta(Y|Y_{obs}))}{\prod_{i \neq j} (1 + \exp(\eta_{ij*}))} \end{aligned}$$

Die Likelihood der *Latent-Space Modelle* mit binärem Response lässt sich in die folgende Form bringen:

$$f(y|\eta) = \frac{\exp(\eta^T y)}{\prod_{i \neq j} (1 + \exp(\eta_{ij}))}$$

Damit ergibt sich eine inhaltliche Interpretation des Minimum-Kullback-Leibler (MKL) Schätzers: Während

man bei der ML-Schätzung die Parameterwerte sucht, die am besten zu den beobachteten Daten y passen, werden bei der MKL-Schätzung diejenigen Parameterwerte gesucht, die am besten zum Erwartungswert der a-posteriori prädiktiven Verteilung $Y|Y_{obs}$, also den vom Modell erwarteten Daten, passen. Damit geht in die MKL-Schätzung im Gegensatz zur ML-Schätzung das komplette statistische Modell inklusive aller Priori Verteilungen mit ein. Für die geschätzten *Latent-Space Modelle* des Publikationsdatensatzes sind in den Abbildungen 3 bis 7 jeweils die MKL-Schätzer der latenten Positionen dargestellt.

5.3.2 Clusterlabels

Die Likelihoods des LCM und des LCREM sind zusätzlich invariant gegenüber Permutationen der Clusterlabels. Bei voller Bayes-Inferenz mit MCMC ergibt sich dabei ein Labelswitching Problem analog zu anderen Mixture-Modellen. Eine mögliche Lösung des Labelswitching Problems ist der Algorithmus nach Stephens (2000) welcher auch bei der Modellschätzung im Paket `latentnet` verwendet wird. Hierbei ergeben sich die Clusterwahrscheinlichkeiten der Knoten durch Minimierung des approximierten a-posteriori erwarteten Verlusts bezüglich der Kullback-Leibler Verlustfunktion, über alle möglichen Permutationen der Clusterlabels (Handcock, Raftery, and Tantrum 2007).

5.3.3 Intercept/Random Effects im LCREM

Im LCREM sind der Intercept und die Random Effects invariant bezüglich additiver Konstanten c :

$$\delta_i^* = \delta_i + c, \forall i$$

$$\beta_0^* = \beta_0 - c$$

Im Paket `latentnet` wird dieses Problem durch die korrelierten Vorschlagsdichten und Block-Updates im verwendeten MCMC-Algorithmus gelöst (Krivitsky et al. 2009).

Quellen

Handcock, Mark S., Adrian E. Raftery, and Jeremy M. Tantrum. 2007. “Model-Based Clustering for Social Networks.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170 (2): 301–54.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. 2002. “Latent Space Approaches to Social Network

Analysis.” *Journal of the American Statistical Association* 97 (460): 1090–8.

Krivitsky, Pavel N., and Mark S. Handcock. 2008. “Fitting Position Latent Cluster Models for Social Networks with Latentnet.” *Journal of Statistical Software* 24 (5): 1–23.

Krivitsky, Pavel N., Mark S. Handcock, Adrian E. Raftery, and Peter D. Hoff. 2009. “Representing Degree Distributions, Clustering, and Homophily in Social Networks with Latent Cluster Random Effects Models.” *Social Networks* 31 (3): 204–13.

Shortreed, Susan, Mark S. Handcock, and Peter D. Hoff. 2006. “Positional Estimation Within a Latent Space Model for Networks.” *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 2 (1): 24.

Stephens, Matthew. 2000. “Dealing with Label Switching in Mixture Models.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62 (4): 795–809.