

Zufall im echten Leben? Wie Zufallsnetzwerke ein romantisches Netzwerk modellieren

Pavel Dimitrov Chachev

Beitrag zur Veranstaltung »Aktuelle Entwicklungen in der Netzwerkforschung« der Sektion Soziologische Netzwerkforschung

Einleitung

Die „Netzwerkmetapher“ (White 2008) hat sich in den letzten Jahrzehnten in der Soziologie nicht nur etabliert, sondern wird flexibel bei einer Vielzahl von empirischen und theoretischen Problemen angewandt. Netzwerke werden unter anderem für die Darstellung von interpersonellen Beziehungen (Wimmer, Lewis 2010), Ko-Investitionen (Podolny 2001), geteilten Bedeutungsmustern (Goldberg 2011), gemeinsames Auftreten in Veranstaltungen (Krüger 2013) aber auch Narrativen (Smith 2007) und Zitationen (Foster et al. 2015) angewandt. Die Liste ist natürlich unvollständig.

So vielfältig die Literatur jedoch auf den ersten Blick erscheinen mag, eine Beobachtung scheinen alle zu teilen: empirische Netzwerke sind nicht zufällig. Netzwerke, die man im „echten Leben“ beobachten kann, haben strukturelle Merkmale, die sie klar von dem klassischen Erdős-Renyi Zufallsmodell (Erdős, Renyi 1959) abgrenzen.

Um einige Beispiele zu nennen: Hillmann und Aven (2011) beobachten Ko-Investitionsnetzwerke im neunzehnten Jahrhundert Russland. Die Struktur ist bei weitem nicht zufällig, da sie für die gegebene Größe unter anderem eine viel zu starke Clusterbildung aufweist, was den Autoren erlaubt, Rückschlüsse über Reputationsmechanismen im Investitionsverhalten zu ziehen. James Moody untersucht in seinem 2001 Papier zu Schulsegregation (Moody 2001) Freundschaftsnetzwerke von Schüler/-innen in den USA. Wieder ist nicht nur die Clusterbildung zu stark, sondern zieht sich nach ganz klaren ethnischen Mustern. Der Autor kommt unter anderem zum Schluss, dass Kontaktmöglichkeiten im schulischen Kontext nicht zur ethnischen Integration beitragen. Als letztes Beispiel sei Bearman und Stovels Studie von narrativen Netzwerken in einer Geschichte genannt. Die untersuchte Geschichte ist von einem Essaywettbewerb, wo Mitglieder der NSDAP schreiben sollten, wie sie Nazis geworden sind. Das narrative Netzwerk weist eine zu starke Fragmentierung auf, um zufällig zu sein. Die Autor/-innen interpretieren die Ergebnisse dahingehend, dass der Effekt einer totalitären Ideologie auf das Konzept des „Selbst“ eines Akteurs in der Netzwerkstruktur zu beobachten ist.

In allen drei Beispielen, seien es Investitionsnetzwerke im neunzehnten Jahrhundert Russland, Freundschaftsnetzwerke in den neunziger Jahren in den USA oder narrative Netzwerke von NSDAP Mitgliedern in Deutschland, die Grundannahmen sind in einer Hinsicht gleich: Netzwerke weisen

strukturelle Merkmale auf, die nicht zufällig sind. Damit sind sie für uns inhaltlich interpretierbar und erlauben uns Rückschlüsse über soziale Mechanismen in der echten Welt zu ziehen.

Gleichzeitig heißt das nicht, dass Netzwerkmodelle deterministisch sind. Modelle mit starken stochastischen Annahmen sind seit Anfang der 2000er populär und haben es geschafft, mit simplen Annahmen typische strukturelle Eigenschaften von Netzwerken zu beschreiben. Zwei der beliebtesten teilweise stochastischen Modelle sind das Watts–Strogatz Modell (Watts, Strogatz 1998) und das Barabasi–Albert Modell (Barabasi, Albert 1999). Watts und Strogatz zeigen mit einem einfachen Neu-Vernetzungs-Algorithmus wie typische Eigenschaften von empirischen Netzwerken, kurze Spannweite und hohes Clustering, modelliert werden können. Barabasi und Albert wiederum demonstrieren wie mit einem Matthäus-Effekt-Algorithmus, erfolgreiche Akteure werden erfolgreicher, die Grad-Verteilung von Netzwerken beschrieben werden können. Die Modelle sind stochastisch aber beinhalten mindestens zum Teil inhaltlich interpretierbare Mechanismen zum Beispiel den Matthäus Effekt. Dabei schaffen sie unter anderem zu zeigen, dass strukturelle Merkmale von empirischen Netzwerken eben nicht zufällig sind.

Im vorliegenden Papier präsentiere ich einen empirischen Fall, der auf das Gegenteil hinweist. Ein romantisches Netzwerk, zum ersten Mal beschrieben von Bearman, Stovel und Moody (2004), weist im Gegensatz zur bisherigen Literatur Eigenschaften auf, die sehr stark einem Erdős–Renyi Netzwerk ähneln und damit gegen die typische sozialwissenschaftliche Intuition gehen.

Das Papier geht folgendermaßen vor: zunächst werden der empirische Fall und die Daten beschrieben. Nach einer kurzen Einführung in Erdős–Renyi Netzwerken werden die Ergebnisse der Simulationen mitsamt Robustheitstests vorgestellt. Am Ende werden Kritikpunkte und Grenzen der Analyse diskutiert.

Der Fall eines romantischen Netzwerks

Als Ausgangspunkt dieser Analyse dient die 2004 im *American Journal of Sociology* erschienene Studie von Bearman, Stovel und Moody (von hier an verkürzt BSM) über ein romantisches Netzwerk in den USA. Die Studie wurde in einer kleinen Stadt in einer durchschnittlichen amerikanischen High School, von den Autor/-innen „Jefferson High“ genannt, durchgeführt. Die jugendlichen Teilnehmende der Studie sind 832 Schüler/-innen der Jefferson High in den Klassen zehn bis zwölf. Für diese Analyse wurden sie gefragt, ob sie innerhalb der letzten 18 Monate eine „besondere romantische Beziehung“¹ (S.56) hatten und sollten ihre drei letzten Partner/-innen inklusive derzeitiger Beziehungen nennen. Teilnehmende, die keine solche Beziehung nennen konnten, wurden des Weiteren gefragt, ob sie „mit jemandem Händchen gehalten haben, jemanden geküsst haben, oder jemanden gesagt haben, dass sie die Person mochten oder geliebt haben“ (S.56). Zusätzlich wurde erhoben, ob die Teilnehmende in nicht-romantischen Beziehungen involviert waren. Dadurch wurden Beziehungen in der Studie miteinbezogen, die nicht romantisch aber dennoch relevant für die Übertragung von sexuellen Krankheiten waren. Die Erhebung wurde im Rahmen der größeren National Longitudinal Study of Adolescent Health (kurz AddHealth) durchgeführt, die ab 1994 über 90.000 Schüler/-innen in 140 Schulen befragt hat².

¹ Übersetzungen aus dem Englischen durch den Verfasser.

² Für alle weiteren Details zur Operationalisierung sei auf das Originalwerk von BSM verwiesen.

Das resultierende Netzwerk hat 573 Knoten, die mindestens eine Beziehung haben und 477 Bindungen. Es wurde vom originalen Papier S.58, Figur 2 händisch abgeschrieben³ und ist in Abbildung 1 dargestellt⁴. Männliche Knoten sind blau gefärbt, weibliche rot.

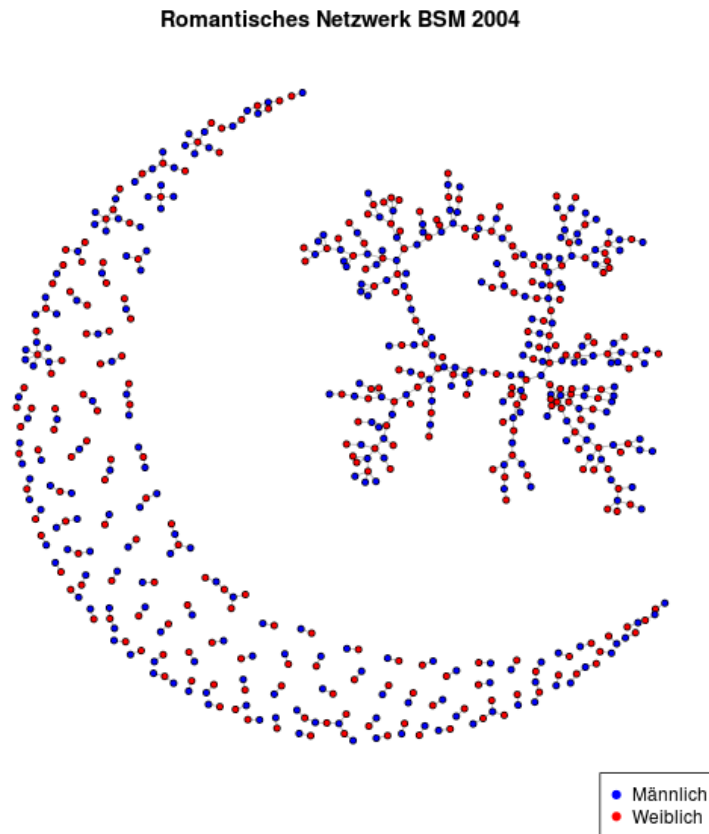


Abbildung1: Romantisches Netzwerk in Jefferson High

Das zentrale Argument von BSM bezieht sich auf die Struktur des Netzwerks. Interventionen und Konzepte, die versuchen, die Ausbreitung von sexuellen Krankheiten zu vermeiden, können je nach Netzwerkstruktur unterschiedliche Effektivität haben. Eine zentralisierte Netzwerkstruktur bedarf anderer Mittel um die Verbreitung von sexuellen Krankheiten vorzubeugen, wie etwa ein Fokus auf stark verbundene Knoten, als ein eher dezentralisiertes Netzwerk wie das in Abbildung 1. Von daher ist es wichtig herauszufinden, inwiefern der konkrete Schulkontext einen Effekt auf die Netzwerkstruktur hat, um potentielle Interventionen antizipieren zu können. Zusätzlich hat man natürlich als Sozialwissenschaftler/-in eine inhärente Motivation, Mikromechanismen hinter emergenten Strukturen aufzudecken, was auch der primäre Anreiz hinter dieser Untersuchung ist.

Eine der wichtigsten Eigenschaften, die das Netzwerk in Jefferson High ausmacht, ist die Abwesenheit von Kreisen der Länge vier. Das Netzwerk kann daher als ein Spannbaum bezeichnet werden,

³ Kleinere Abweichungen in der Gradverteilung können auftauchen. Der Knotengrade zwei, drei und vier kommen original 139, 59 und 28 Mal vor, hier jedoch 140, 57 und 30. Die Ergebnisse der Analyse sollten im Kern nicht davon betroffen sein. Neben der originalen Darstellung dient folgende Quelle: <http://faculty.washington.edu/stovel/adolescents.shtml> (Zugegriffen: 15.09.2019).

⁴ Wenn fortan vom „empirischen Netzwerk in Jefferson High“ gesprochen wird, ist selbstverständlich nur die Rekonstruktion gemeint.

denn es enthält gerade mal drei solche Kreise. BSM implementieren entsprechend ein Modell, was Heterosexualität, Homophilie, und am wichtigsten, ein Verbot der Kreise der Länge vier enthält. Auf diese Weise schaffen sie es, essentielle strukturelle Eigenschaften des Netzwerks zu reproduzieren. Das Verbot der Kreise der Länge vier ist substantiell als ein Verbot für das „daten“ des Ex-partners des Partners deines Ex-partners zu verstehen.

Simulationen und Ergebnisse

Das Simulationsmodell

Um die Struktur des romantischen Netzwerks zu simulieren, wird hier ein anderer Ansatz verwendet. Ähnlich wie bei BSM ist das Ziel, mit einem Modell die Netzwerkstruktur möglichst genau zu reproduzieren. Diese Art der Erklärung ist bekannt als „generative Erklärung“ (Epstein 2006): Wenn man ein Phänomen nicht generieren kann, dann kann man es nicht erklären.

Das hier verwendete Modell ist jedoch ein sogenanntes Erdős–Renyi Netzwerk (Erdős, Renyi 1959). Als Input nimmt die hier verwendete $G(n,m)$ Version eine Summe Knoten und eine Summe Kanten und verteilt die Kanten zufällig an Paare von Knoten.

Die Simulationen wurden in R Version 3.4.4 mithilfe des `igraph` 1.0.1 Pakets durchgeführt. Der gesamte Code nötig für die Simulation sieht folgendermaßen aus:

```
library(igraph)
er_list_2<-list()
set.seed(1)
for(i in 1:1000) {
  er_list_2[[i]]<-erdos.renyi.game(n=832,p.or.m=477,type="gnm", directed =
  FALSE)
}
```

Das Modell wird anhand der empirischen Kanten- und Knotenmengen des Netzwerks von Jefferson High kalibriert. Das heißt es werden Zufallsnetzwerke mit 832 Knoten und 477 Kanten modelliert. Die Messparameter wurden zum Zweck der Vergleichbarkeit mit der vorherigen Literatur möglichst ähnlich zu den Parametern des Originalwerks von BSM gehalten.

Es wird die Bonacich-Power Zentralität (Bonacich 1987) mit einem Exponenten von -0.7 benutzt. Zusätzlich wird auch die Summe der Triaden gemessen, um zu überprüfen, ob das Modell eventuell mehr homosexuelle Beziehung erschafft als man anhand der Daten erwarten würde.

Ergebnisse

Abbildung 2 zeigt die Ergebnisse von 1000 simulierten Erdős–Renyi Netzwerken kalibriert über die Werte von Jefferson High. Rote Punkte markieren die tatsächlichen empirischen Werte. Die Werte wurden anhand einer Standardnormalverteilung standardisiert.

Mehrere Beobachtungen ergeben sich aus Abbildung 2: Erstens kann man feststellen, dass die empirischen Werte auf mehreren Parametern erfolgreich modelliert werden. Werte des tatsächlichen Netzwerks bezüglich der Summe der Isolate sowie der Durchmesserlänge sind beide innerhalb des Interquartilsabstandes der Simulationen.

Die Zentralität wird zudem exakt modelliert: der simulierte Zentralitätswert ist 0.0287 verglichen mit einem empirischen Wert von 0.0283.

Der strukturell relevanteste Parameter für das Netzwerk wird präzise getroffen: die Summe der Kreise der Länge vier. In Abbildung 2 ist der standardisierte empirische Wert zwar viel höher als der Empirische, was jedoch daran liegt, dass es kaum Abweichung in den Simulationen gibt.

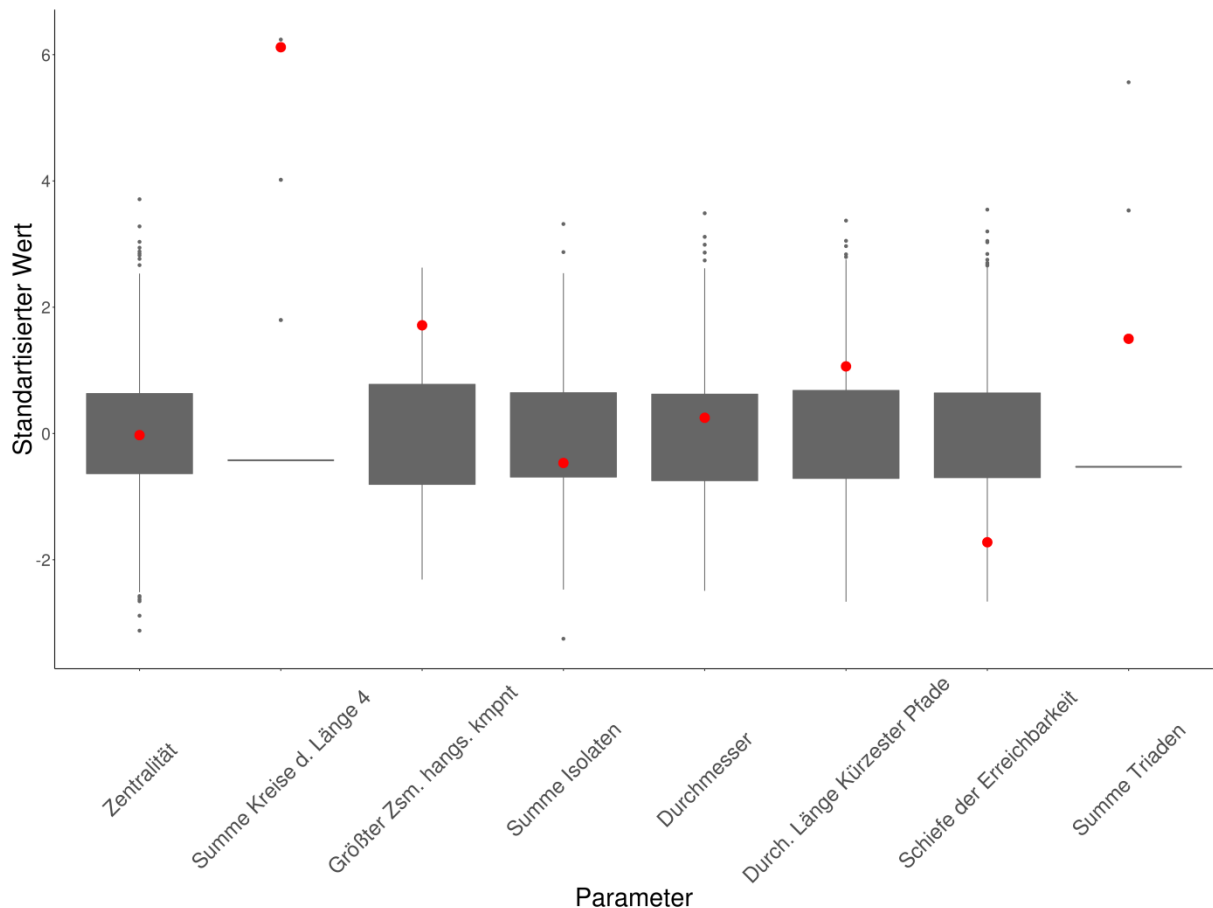


Abbildung 2: Ergebnisse von 1000 simulierten Zufallsnetzwerken

$|V|=832$, $|E|=477$. Rote Punkte zeigen die empirischen Werte

Die beobachtete Summe von Kreisen ist drei. Die Simulationen ergeben im Durchschnitt null. Relativ zur Netzwerkgröße ist die Aussage dann richtig, dass die Simulationen bezüglich dieses Parameters das empirische Netzwerk erfolgreich modellieren.

Die Abwesenheit von Kreisen der Länge drei zeigt dazu, dass ein potentieller Nachteil von Erdős-Renyi Netzwerken in dieser Spanne von Werten noch nicht zum Tragen kommt: die fehlende Unterscheidung nach Geschlecht. Es mag zunächst wie ein starker Nachteil wirken, dass ein Modell für die Entstehung sexueller Beziehungen das Geschlecht komplett ignoriert. Strukturell scheint das aber keinen Unterschied zu machen.

Es wäre nämlich problematisch, wenn das Modell zu viele Kreise der Länge drei produzieren würde, denn das würde klar auf eine Überschätzung der Summe homo- und bisexuellen Beziehungen

hinweisen⁵. Da aber ähnlich wie die Kreise der Länge vier auch die Kreise der Länge drei recht genau geschätzt werden, der empirische Wert ist eins, der simulierte Mittelwert ist null, ist das kein Problem.

Was aber durchaus problematisch ist, sind die restlichen simulierten Werte. Die Größe des größten Zusammenhangskomponenten, die Schiefe der Erreichbarkeitsmatrix sowie die durchschnittliche Länge der kürzesten Pfade weisen alle auf ein Problem hin: die Größe des größten Zusammenhangskomponenten ist zu klein.

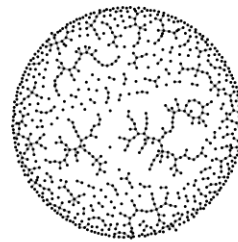
Die durchschnittliche simulierte Größe ist dabei 182 Knoten, die Teil des größten Zusammenhangskomponenten sind. Verglichen mit einer empirischen Größe von 288 ist das ein durchaus eklatanter Unterschied. Eine visuelle Inspektion bestätigt die ursprüngliche Intuition.

Abbildung 3 zeigt mehrere unterschiedliche Simulationsdurchläufe. Spannbaumähnliche Strukturen werden erfolgreich in allen Fällen produziert. Jedoch kann es passieren, dass das Netzwerk zu fragmentiert ist, wie man an Durchlauf 151 und 641 beobachten kann. Abbildungen 4 und 5 erlauben den direkten Vergleich mit dem empirischen Netzwerk. Hier wird das Muster noch deutlicher: Während einige Durchläufe das empirische Netzwerk fast genau reproduzieren, schaffen es andere Durchläufe nicht einmal, einen wirklichen größten Zusammenhangskomponenten zu erschaffen.

Simulation Durchgang 151



Simulation Durchgang 641



Simulation Durchgang 395



Simulation Durchgang 596

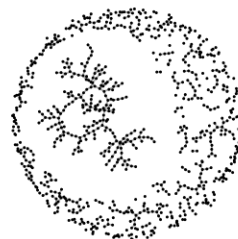


Abbildung 3: Beispiel Simulationsdurchläufe

⁵ Vielen Dank an Haiko Lietz für diesen Hinweis.

Simulation Durchgang 596

Romantisches Netzwerk BSM 2004

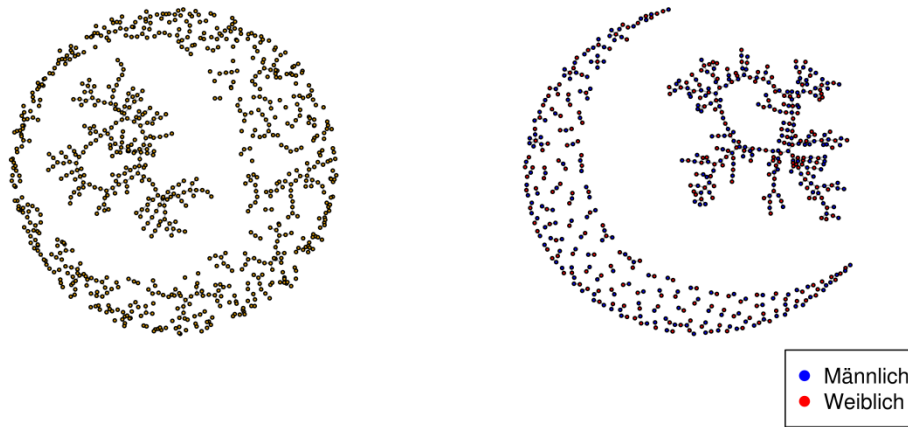


Abbildung 4: Simulationsdurchlauf 596 und das empirische Netzwerk von Jefferson High

Simulation Durchgang 151

Romantisches Netzwerk BSM 2004

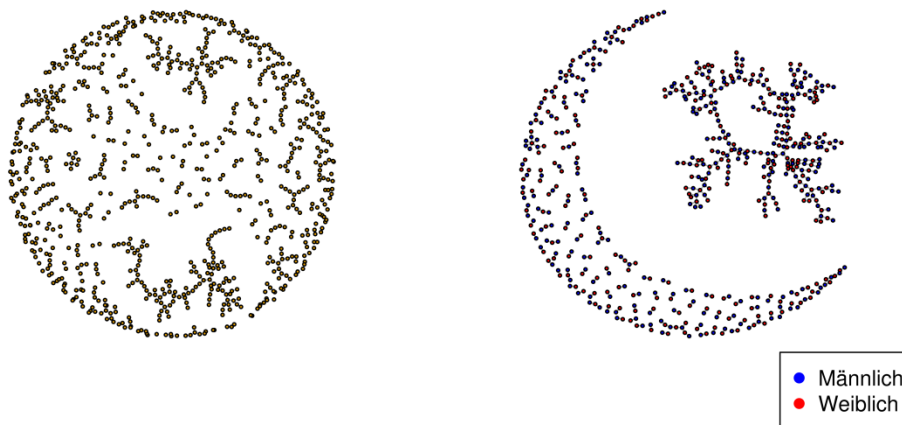


Abbildung 5: Simulationsdurchlauf 151 und das empirische Netzwerk von Jefferson High

Robustheit

Um zu überprüfen, inwiefern das Modell robust ist und in welcher Reichweite von Parametern eventuell sogar bessere Ergebnisse möglich sind, werden nun Robustheitsüberprüfungen durchgeführt. Es ist soweit klar, dass Probleme bei der Modellierung durch die Dynamik des größten Zusammenhangskomponenten verursacht werden. Gleichzeitig ist bekannt, dass die Wahrscheinlichkeit einen größten Zusammenhangskomponenten zu bekommen in einem Erdős–Renyi Modell von der Dichte abhängt (Newman 2003).

Im Folgenden wird deswegen durch Variation der Knotenmenge die Dichte des Netzwerks verändert, um für relevante Messparameter die Robustheit des Modells zu überprüfen. Das heißt es werden Isolate hinzugefügt oder entfernt.

Die Vorgehensweise hat auch eine substantielle Interpretation. Wir beobachten 832 Schüler/-innen in Jefferson High. Jedoch ist unklar, ob alle tatsächlich für die sexuelle Interaktion relevant waren. Es gab Beziehungen außerhalb der Schule und auch wenn es eine sinnvolle Annahme ist, muss sexuelles Interesse nicht bei allen Akteuren vorhanden sein. Tatsächlich ist es keine triviale Frage. BSM haben sich zum Beispiel entschieden, nur die 573 Akteure in die Simulation miteinzubeziehen, die mindestens eine Bindung hatten. Das wiederum ist auch kritisch, da unklar ist, ob nicht eventuell mehr Akteure Interesse an Beziehungen hatten, diese aber nie entstanden sind. Der Robustheitstest ist damit als ein „Alternativszenario“ zu denken: wie würde Parameter X gemäß der Simulation aussehen, wenn zum Beispiel 730 Leute in der Schule an sexuellen oder romantischen Kontakten interessiert wären (wobei wir wissen dass am Ende 477 Bindungen entstehen).

Abbildung 6 zeigt die Ergebnisse für den Robustheitstest für die Größe des größten Zusammenhangskomponenten. Auf der x-Achse ist die jeweilige Knotenmenge abgebildet, auf der y-Achse die absolute Größe des größten Zusammenhangskomponenten. Die Boxplots zeigen die Ergebnisse von 100 Erdős–Renyi Netzwerken mit 477 Bindungen und der jeweiligen Knotenanzahl. Rote Punkte repräsentieren den empirischen Wert. Die roten Linien repräsentieren die jeweils möglichen minimalen und maximalen Knotenmengen. So ist es für die Interpretation zum Beispiel nicht sinnvoll, weniger als 573 oder mehr als 832 Akteure in dem Netzwerk zu haben. Ein „Alternativszenario“ mit 300 Leuten zum Beispiel würde keinen Sinn machen, da wir wissen, dass mindestens 573 Akteure in dem Netzwerk sind. Dennoch wurden die Wertebereiche hinzugefügt, um die Dynamik der Simulation besser abschätzen zu können.

Der blaue Bereich entspricht dem Bereich, wo die empirischen Werte zumindest im Interquartilsabstand der Simulationen sind.

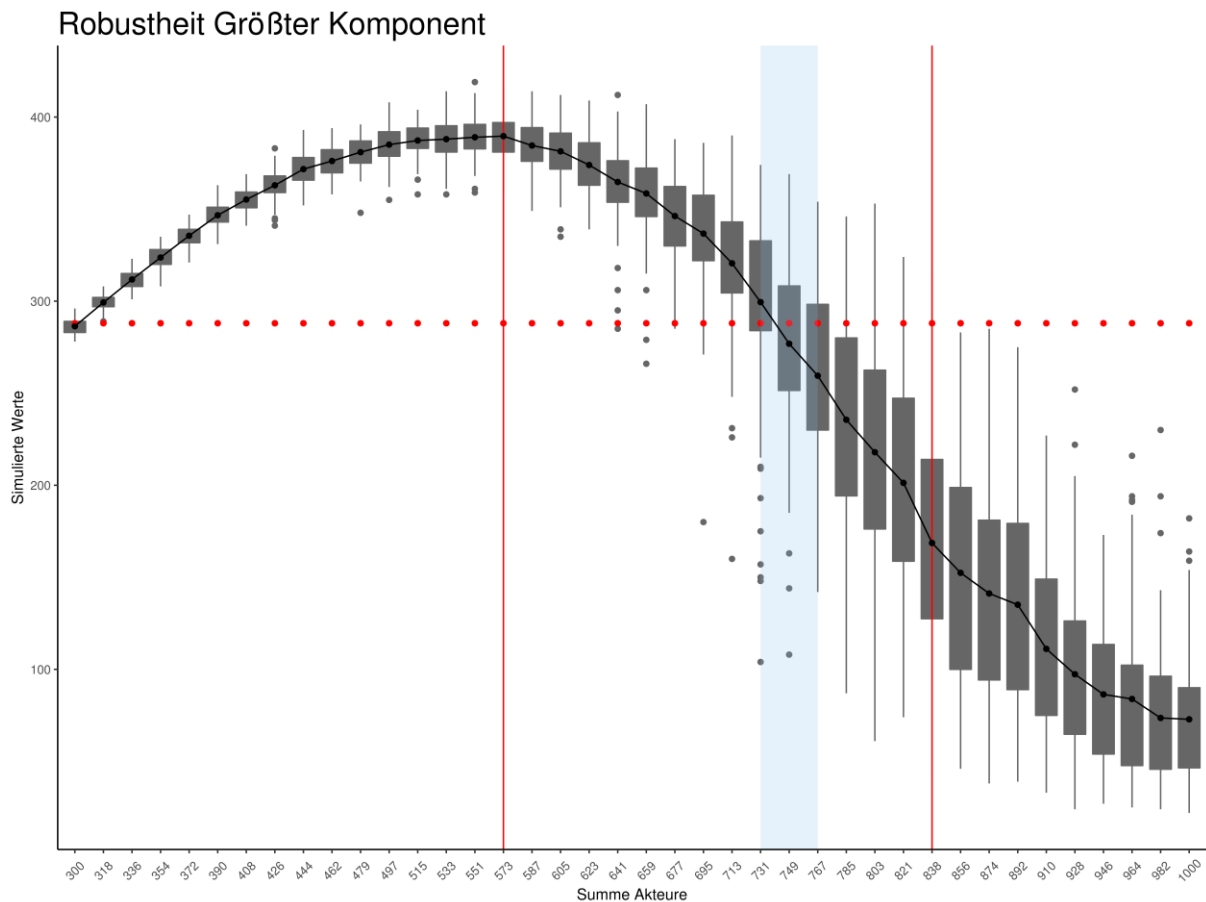


Abbildung 6: Robustheitstest für die Größe des größten Zusammenhangskomponenten

Der Robustheitstest zeigt, dass Erdős-Rényi Netzwerke keine sehr robusten Ergebnisse bezüglich der Größe des größten Zusammenhangskomponenten liefern. Es ist zwar möglich, den empirischen Wert bei ungefähr 750 angenommenen Akteuren exakt zu modellieren. Doch der Bereich, in dem akzeptable Werte produziert werden, ist gerade mal in einer Reichweite von 36 Akteuren. Das bedeutet zwar, dass zwischen 731 und 767 Akteuren die Simulation das empirische Netzwerk modelliert. Da das jedoch kein großer Bereich ist und man bei 832 angenommenen Akteuren den Parameter schon um 100 daneben schätzt, ist das Modell bezüglich der Größe des größten Komponenten nicht wirklich robust. Das ist wenig überraschend, da eine der wichtigen Eigenschaften von Erdős-Rényi Graphen der plötzliche Anstieg, „double jump“, in der Größe des größten Zusammenhangskomponenten ist (Newman 2003, S.198).

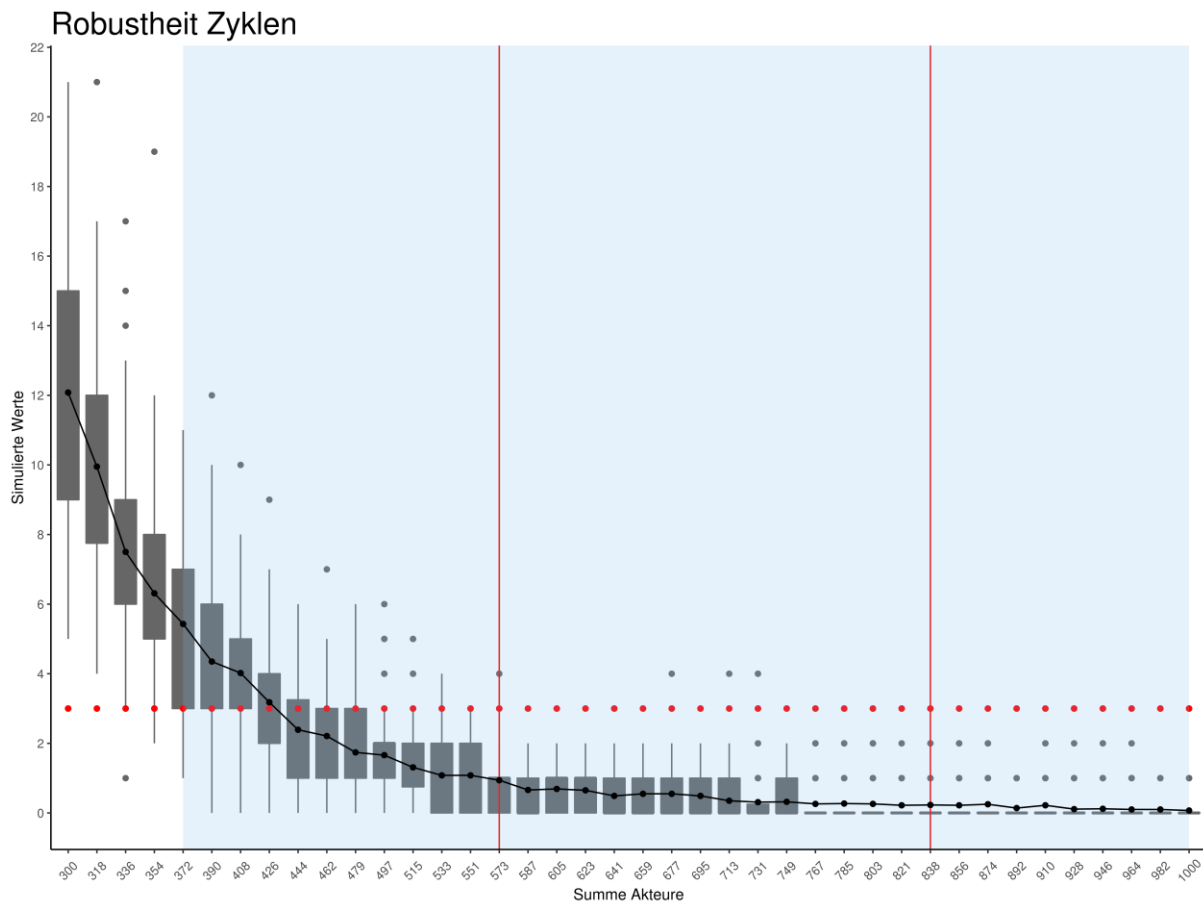


Abbildung 7: Robustheitstest für die Summe der Kreise der Länge vier

Andere strukturelle Maße sind jedoch weitaus robuster. Abbildung 7 zeigt den gleichen Robustheitstest jedoch für die Summe der Kreise der Länge vier. Offensichtlich wird diese Eigenschaft erfolgreich und robust simuliert. Spätestens ab 426 Akteuren kann man sagen, dass der Parameter erfolgreich simuliert wird. Ab 573 Akteuren produzieren Erdős-Renyi Netzwerke zuverlässig Spannbaumstrukturen.

Als letztes zeigt Abbildung 8 die Robustheitsanalyse für die Bonacich-Zentralität. Zwar ist die Simulation nicht so robust bezüglich Zentralität wie bezüglich der Summe der Kreise der Länge vier. Dennoch werden die Werte erfolgreich in einem Bereich zwischen 695 und 856 potentiellen Akteuren modelliert, immerhin ein Bereich von ungefähr 150 Akteuren. Zudem ist die Steigung der simulierten Mittelwerte abhängig von der Knotenmenge graduell. Im Gegensatz zur Größe des größten Zusammenhangskomponenten verursachen Erdős-Renyi Netzwerke also keine „double jumps“ und sind damit ein zuverlässigeres Modell was Zentralität angeht.

Alles in einem bestätigen die Robustheitstests die ursprünglichen Beobachtungen. Erdős-Renyi Netzwerke reproduzieren erfolgreich und robust wichtige strukturelle Eigenschaften wie etwa die Abwesenheit von Kreisen der Länge vier oder die Zentralität des Netzwerkes. Sie sind jedoch nicht robust bezüglich der Größe des größten Zusammenhangskomponenten.

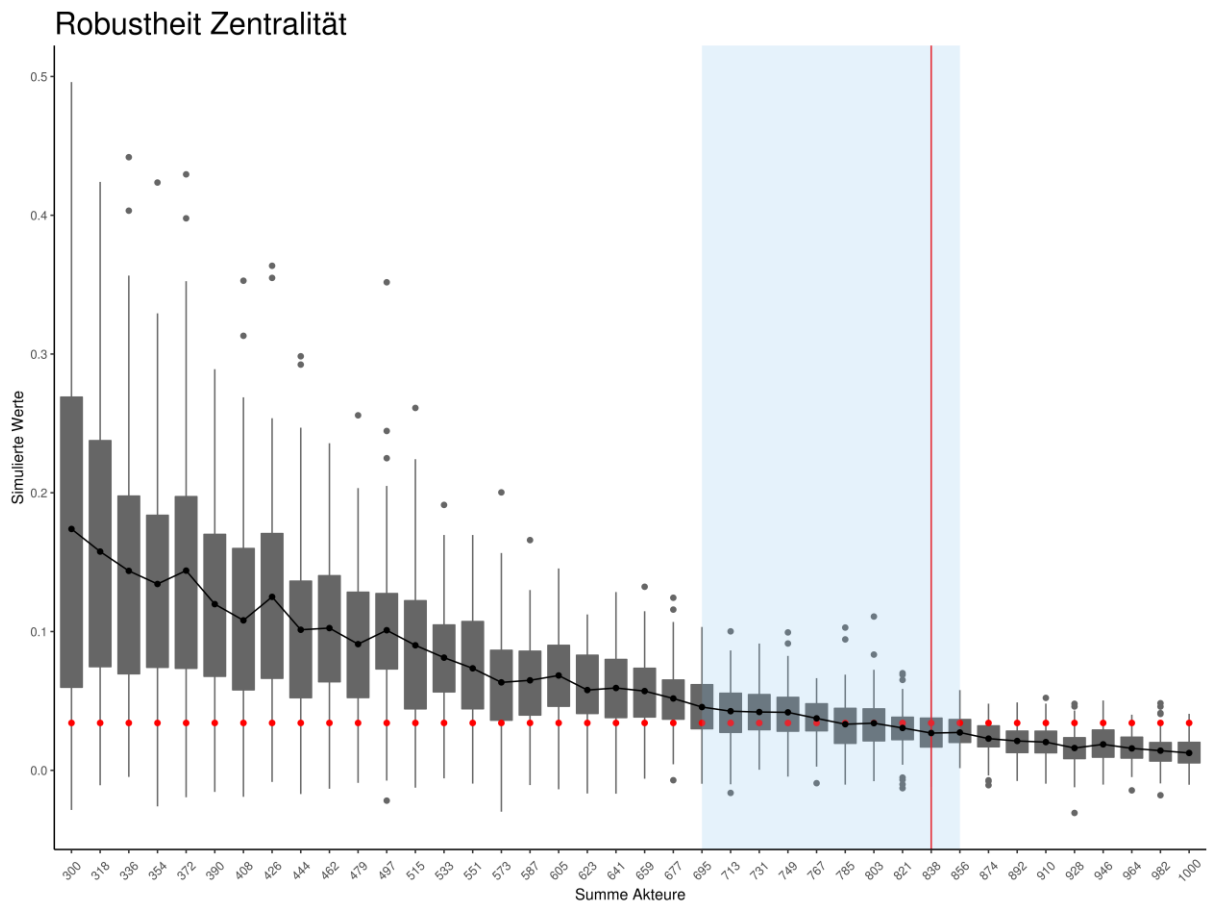


Abbildung 8: Robustheitstest für die Zentralität

Diskussion

Im vorliegenden Papier wurde der Fall eines romantischen Netzwerks in einer amerikanischen High School präsentiert. Es wurde versucht zu überprüfen, inwiefern einfache Zufallsnetzwerke in der Lage sind, die Netzwerkstruktur in Jefferson High zu modellieren. Die Ergebnisse gehen stark gegen jegliche sozialwissenschaftliche Intuition, indem sie zeigen, dass mit extrem simplen Annahmen, der Knoten- und Kantenmenge, einige strukturelle Parameter robust und genau modelliert werden können. Einzig die Größe des größten Zusammenhangskomponenten ist kein robuster Parameter auch wenn es unter Lockerung der angenommenen Summe der Akteure möglich ist auch ihn in einem begrenzten Bereich zu reproduzieren.

Trotz dieses Nachteils ist das Ergebnis dennoch überraschend, da die meisten, wenn nicht alle, sozialwissenschaftlichen Modelle auf der Annahme ruhen, dass Akteure sinnhaft handeln und auch aggregierte Strukturen für Forscher/-innen sinnhaft interpretierbar sind, eben weil sie nicht zufällig sind. Dazu kommt, dass es sich hierbei um romantische und sexuelle Beziehungen in einem Schulkontext handelt: einem Bereich, wo man am wenigsten vermuten würde, dass Regeln und Normen das Handeln der Akteure leiten.

Eine der seltenen Ausnahmen diesbezüglich ist die ethnografische Studie von Kibanov et al. (2019), die Bewegungsmuster von Wissenschaftler/-innen bei einem „PhD retreat“ mit RFID-tags gemessen haben und zufällige Kontaktmuster beobachten konnten.

Interessanterweise haben Jackson und Rogers sich schon 2007 gefragt wie zufällig soziale Netzwerke überhaupt sind. Sie analysieren sechs verschiedene soziale Netzwerke, inklusive des Falles von BSM. Dabei kommen sie auch zum Schluss dass dieser Fall fast einer uniformen Zufallsverteilung folgt („almost uniformly random“). Die Autoren können sogar noch darüber hinaus mit wenigen Anpassungen auch Eigenschaften weiterer Netzwerke modellieren⁶.

Entsprechend ist ein kritisches Hinterfragen der Modellannahmen hier besonders angebracht. Zunächst einmal bezieht das Modell Geschlecht nicht mit ein. Akteure erschaffen Beziehungen ohne nach irgendwelchen Eigenschaften zu unterscheiden, was keine realistische Annahme ist. Dazu kommt, dass eben weil es bisher keine anderen, dem Autor bekannten, erfolgreichen empirischen Simulationen mit Hilfe von Erdős–Renyi Netzwerken gibt, man eher Grund zum Zweifeln hat, ob das Modell hier wirklich passend geschweige denn endgültig ist.

Als letzter Kritikpunkt bleibt der Vergleich zur bisherigen Literatur. Spannbäume sind zunächst selten zu beobachten und Netzwerke, wo das der Fall ist, sind eventuell schwer zu vergleichen. Potterat et al, (2002) untersuchen in ihrer Studie bi- und homosexuelle Beziehungen in Colorado Springs. Jedoch geht es um HIV infizierte Fälle. Die Netzwerkstruktur ähnelt zudem eher dem Barabasi–Albert Model als dem Spannbaum bei BSM. Cacioppo, Christiakakis und Fowler (2009) zeigen als Teil der Framington Heart Study ein Netzwerk, das eine Spannbaumstruktur aufweist. Jedoch handelt es sich hierbei um Freundschafts-, Geschwister- und Ehebindungen. Von daher ist es fraglich, inwiefern es vergleichbar ist⁷.

Der vergleichbarste Fall wurde von Hellerigner und Kohler (2007) berichtet. Die Autoren haben auf Likoma, einer Insel in Malawi, alle sexuellen Beziehungen von Personen zwischen 18 und 35 erhoben. Das dort berichtete Netzwerk würde von dem hier verwendeten Modell nicht unterstützt, da bei der berichteten Dichte von 2043 Bindungen zu 923 Knoten man nach dem Erdős–Renyi Modell keine Spannbäume mehr erwarten würde, was aber in Likoma zu beobachten ist. Dazu muss gesagt werden, dass im AddHealth nur Bindungen innerhalb von Jefferson High erhoben wurden. Damit weiß man nicht wie das Gesamtnetzwerk des Ortes aussieht, womit es schwer mit Likoma zu vergleichen ist. Jedoch ist klar, dass Likoma der soweit vergleichbarste Fall ist, wo jedoch ein Transfer des Modells nicht gelingen würde.

Andersherum versucht die zitierte Literatur selten wirklich Netzwerkstrukturen generativ zu modellieren. BSM ist einer der wenigen konsequenten Versuche, weswegen es auch diese Arbeit hier inspiriert hat⁸. Das hier präsentierte Modell hat aber weniger Annahmen, ist einfacher und liefert bessere Ergebnisse. Auch wenn es nicht immer robust oder realistisch ist, sollte man deswegen nicht ausschließen, dass es ein guter Ausgangspunkt für die Analyse von Spannbäumen ist und man mit wenigen Anpassungen in zukünftigen Arbeiten es auch auf Kontexten außerhalb von Jefferson High anwenden kann.

⁶ Vielen Dank an Florian Andersen der mich auf die großartige Arbeit von Jackson und Rogers, zweier Ökonomen, hingewiesen hat. Einerseits bestätigen ihre Ergebnisse die hier vorgelegte Analyse, auch wenn deren Ansatz leicht anders ist. Andererseits ist die Tatsache, dass ich erst ein Jahr nach dem DGS Kongress zufälligerweise von deren Arbeit erfahren habe, ein Indiz für den tief enttäuschenden Zustand von interdisziplinärer Kommunikation.

⁷ Man kann hier natürlich auch argumentieren, dass Erdős–Renyi Netzwerke, eben weil sie kaum Annahmen treffen, leicht auf unterschiedliche Kontexte übertragbar sind.

⁸ An dieser Stelle sei James Moody gedankt, der extrem schnell und transparent jegliche Nachfragen zu der Modellierung in dem Papier beantwortet hat. Vielen Dank auch an Anna Kaiser für das Feedback und Unterstützung.

Literatur

- Barabasi, A-L. und A Reka. 1999. Emergence of Scaling in Random Networks. *Science* 286:509–512.
- Bearman, P. S., J. Moody und K. Stovel. 2004. Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks. *American Journal of Sociology* 110(1):44–91.
- Bonacich, P. 1987: Power and Centrality: A Family of Measures. *American Journal Of Sociology* 92(5):1170–1182.
- Cacioppo, J. T., J. H. Fowler und N. A. Christakis. 2009. Alone in the Crowd: The Structure and Spread of Loneliness in a Large Network. *Journal of Personality and Social Psychology* 97(6):977–991.
- Csardi, G. und T. Nepusz. 2006. The igraph software package for complex network research, Interjournal, Complex Systems 1695. URL <http://igraph.org>
- Epstein, J. M. 2006. *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press.
- Erdős, P. und A. Renyi. 1959. *On Random Graphs I*. Publicationes Mathematicae.
- Foster, J. G., A. Rzhetsky und J. A. Evans. 2015. Tradition and Innovation in Scientists' Research Strategies. *American Sociological Review* 80:875–908.
- Goldberg, A. 2011. Mapping Shared Understandings Using Relational Class Analysis: The Case of the Cultural Omnivore Reexamined. *American Journal of Sociology* 116(5):1397–1436.
- Helleringer, S. und H.-P. Kohler. 2007. Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi. *AIDS* 21:2323–2332.
- Jackson, M. O. und B.W. Rogers. 2007. Meeting Strangers and Friends of Friends: How Random Are Social Networks? *The American Economic Review* 97(3):890–915.
- Kibanov, M, R. H. Heiberger, S. Rödder, M. Atzmüller und G. Stumme. 2019. Social studies of scholarly life with sensor-based ethnographic observations. *Scientometrics* 119(2):1387–1428.
- Krüger, U. 2013. *Meinungsmacht*. Herbert von Harlem Verlag.
- Moody, J. 2001. Race, School Integration and Friendship Segregation in America. *American Journal of Sociology* 107(3):679–716.
- Newman, M. E. J. 2003. The Structure and Function of Complex Networks. *SIAM Review* 45(2):167–256.
- Podolny, J. M. 2001. Networks as the Pipes and Prisms of the Market. *American Journal of Sociology* 107(1):33–60.
- Potterat, J. J., L. Philipps-Plummer, S. Q. Muth, R. B. Rothenberg, D. E. Woodhouse, T. S. Maldonado-Long, H. P. Zimmerman und J. B. Muth. 2002. Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs. *Sexually Transmitted Infections* 78(Supplement 1):i159–i163.
- R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Smith, T. 2007. Narrative boundaries and the dynamics of ethnic conflict and conciliation. *Poetics* 35:22–46.
- Watts, D. J. und S. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature* 393:440–442.
- White, H. C. 2008: *Identity and Control. How Social Formations Emerge*. 2nd Edition. Princeton University Press.
- Wimmer, A. und K. Lewis. 2010. Beyond and Below Racial Homophily: ERG Models of Friendship Network Documented on Facebook. *American Journal of Sociology* 116(2):583–642.