

Universität Konstanz  
Fachbereich Informatik und Informationswissenschaft  
Lehrstuhl für Algorithmenik  
Wintersemester 2009/2010  
Algorithmenik-Seminar Netzwerkmodelle  
Betreuer 1: Prof. Dr. Ulrik Brandes  
Betreuer 2: Uwe Nagel

# **Das Social Space Modell**

**Models of social networks based on social distance attachment**

**Analyse der Arbeit von  
Boguñá, Pastor-Satorras, Díaz-Guilera und Arenas**

Thomas Ummenhofer  
thomas.ummenhofer@uni-konstanz.de  
Eschenweg 3, 78479 Reichenau  
4. April 2010

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>3</b>
1.1	Hintergrund . . . . .	3
1.2	Modellierung sozialer Netzwerke . . . . .	3
<b>2</b>	<b>Charakteristiken sozialer Netzwerke</b>	<b>3</b>
2.1	Das PGP Web of Trust . . . . .	3
2.1.1	Kritik an der Verwendung des PGP Web of Trust . . . . .	4
2.2	Netzwerkanalyse . . . . .	4
2.2.1	Clustering Coefficient . . . . .	4
2.2.2	Positive Gradkorrelation . . . . .	5
2.2.3	Community Structure . . . . .	6
2.2.4	Kritik . . . . .	7
<b>3</b>	<b>Das Social Space Modell</b>	<b>8</b>
3.1	Raum . . . . .	8
3.2	Abstraktion auf das Modell . . . . .	9
3.3	Verbindungswahrscheinlichkeit . . . . .	9
3.4	Kritik am Aufbau des Social Space Modells . . . . .	10
<b>4</b>	<b>Analyse der erzeugten Netzwerke</b>	<b>10</b>
4.1	Clustering Coefficient . . . . .	10
4.2	Korrelation des Knotengrades . . . . .	12
4.3	Community Structure . . . . .	13
4.4	Kritik an der Netzwerkanalyse . . . . .	13
4.5	Verwendung im Attract & Introduce Modell . . . . .	14
<b>5</b>	<b>Fazit</b>	<b>15</b>

# 1 Einleitung

## 1.1 Hintergrund

Diese Arbeit entstand im Seminar "Netzwerkmodelle", welches das gleichnamige Projekt begleitet. Ziel des Projekts ist es, die Ergebnisse der Veröffentlichung "Model of genetic variation in human social networks" [JHF09], welche sich mit der Modellierung sozialer Netzwerke beschäftigt, zu reproduzieren und kritisch zu bewerten. Die Autoren dieser Veröffentlichung verglichen ihre Ergebnisse mit denen einiger anderer Verfahren zur Netzwerkmodellierung.

Diese Seminararbeit beschäftigt sich hauptsächlich mit einem dieser Modelle, dem Social Space Modell [MBn04]. Mein Ziel ist es nicht nur, dieses Modell zu erklären, sondern auch die Vorgehensweise der Autoren nachzuvollziehen und kritisch zu hinterfragen.

## 1.2 Modellierung sozialer Netzwerke

Soziale Netzwerke zeigen die Komplexität menschlicher Interaktion. Sie können durch Graphen dargestellt werden, bei denen die Individuen durch Knoten und die Beziehungen zwischen den Individuen durch Kanten repräsentiert werden. Die Darstellung als Graph erlaubt die statistische Charakterisierung der Struktur sozialer Netzwerke. Soziale Netzwerke weisen ganz bestimmte Charakteristiken auf, weshalb es ganz bestimmte Mechanismen geben muss, welche ihre Bildung beeinflussen. Viele Soziologen haben es sich zur Aufgabe gemacht, Modelle vorzuschlagen, welche die Bildung sozialer Netzwerke beschreiben. Das Ziel ist es, ein mathematisches Modell zu gewinnen, welches Netzwerke mit der selben Struktur erzeugt, wie sie auch reale soziale Netzwerke aufweisen.

# 2 Charakteristiken sozialer Netzwerke

## 2.1 Das PGP Web of Trust

Bevor ein Modell zur Erzeugung sozialer Netzwerke vorgestellt wird, muss untersucht werden, welche Strukturen charakteristisch für soziale Netzwerke sind. Dazu führt man verschiedene Netzwerkanalysen auf real existierenden sozialen Netzwerken durch. In diesem Fall wurde dafür das PGP Web of Trust [?] verwendet.

PGP ist ein weit verbreitetes Verschlüsselungsverfahren. Sein Vorteil besteht darin, dass zum Verschlüsseln und zum Entschlüsseln verschiedene Schlüssel verwendet werden. Der Public Key zum verschlüsseln kann jedem bekannt sein. Die Nachricht

entschlüsseln kann man aber nur, wenn man den Privat Key kennt. Will man einer Person eine verschlüsselte Nachricht schicken, fragt man sie nach ihrem Public Key und verschlüsselt damit die Nachricht. So wird sichergestellt, dass nur die Person, die den entsprechenden Privat Key besitzt, die Nachricht lesen kann. Diese Vorgehensweise hat aber den gewaltigen Nachteil, dass der Absender der Nachricht sich niemals 100% sicher sein kann, dass der benutzte Public Key tatsächlich zu der Person gehört, an die man die Nachricht schicken will. Genau hier setzt das PGP Web of Trust an. Teilnehmer des Netzwerkes können sich gegenseitig ihre Public Keys signieren, wenn sie sich absolut sicher sind, dass sich hinter dem Schlüssel auch die Person verbirgt, die sie vorgibt zu sein. Auf diese Weise entsteht ein gerichteter Graph, bei dem die Knoten die Teilnehmer repräsentieren und die Kanten geben an, wer wessen Schlüssel signiert hat.

Im Juli 2001 hatte dieses Netzwerk 191548 Knoten und 286290 Kanten. Dieses Netzwerk wurde noch weiter vereinfacht, so wurden für die Analyse nur Kanten zugelassen, zu denen auch eine Rückrichtung existiert (57243 Knoten, Grad  $\langle k \rangle = 2.16$ ). Außerdem wurde die größte Zusammenhangskomponente dieses Netzwerkes nochmal getrennt betrachtet (10680 Knoten, Durchschnittsgrad  $\langle k \rangle = 4.55$ ).

### **2.1.1 Kritik an der Verwendung des PGP Web of Trust**

Es wird leider nicht klar, wieso ausgerechnet das PGP Web of Trust als Beispiel für ein reales soziales Netzwerk verwendet wird. Es wird nicht darauf eingegangen, was ein soziales Netzwerk ausmacht und wieso das PGP Web of Trust ein geeigneter Repräsentant ist. Außerdem gibt es einige andere reale Netzwerke, deren Daten auch verfügbar sind, die aber eindeutig der Gruppe der sozialen Netzwerke angehören (z.B. Add Health [Add]).

## **2.2 Netzwerkanalyse**

Es wurden 3 verschiedene Analysen des PGP Web of Trust vorgenommen. Die vom Social Space Modell erzeugten Netzwerke werden später mit den selben Verfahren analysiert in der Hoffnung, ähnliche Ergebnisse zu bekommen.

### **2.2.1 Clustering Coefficient**

Der Clustering Coefficient misst die Transitivität der Verbindungen. Er quantifiziert die Vorstellung, dass jedes Individuum Mitglied verschiedener Gruppen ist, innerhalb welcher sich alle Individuen kennen.

Der Clustering Coefficient  $c$  eines Knotens  $i$  entspricht dem Verhältnis der existie-

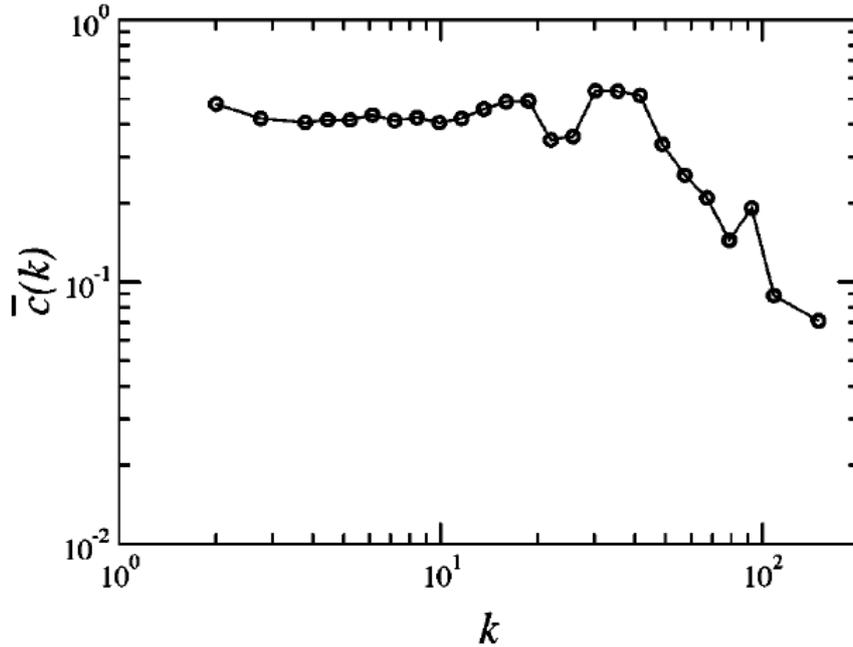


Abbildung 1: Durchschnittlicher Clustering Coefficient im PGP Web of Trust in Abhängigkeit zum Grad

renden Dreiecke  $e_i$ , welche den Knoten enthalten, zu der Anzahl der Dreiecke die möglich sind, wenn alle  $k$  Nachbarn von  $i$  kombiniert werden.

$$c_i = \frac{e_i}{\frac{1}{2}k_i(k_i - 1)}$$

Den Clustering Coefficient  $\langle c \rangle$  des gesamten Netzwerkes entspricht dem durchschnittlichen Clustering Coefficient aller Knoten. Der tatsächliche Wert von  $\langle c \rangle$  kann natürlich von Netzwerk zu Netzwerk variieren. Wichtig für soziale Netzwerke sind aber 2 Punkte. Erstens muss  $\langle c \rangle$  einen Wert haben, der höher ist als bei zufälliger Verbindung mancher Knoten. Im PGP Web of Trust ist  $\langle c \rangle = 0.4$ . Außerdem kann man die Verteilung von  $c$  in Abhängigkeit zum Grad betrachten. Wie Abbildung 1 zeigt, ist dieser Wert für relevante Knotengrade konstant. Daher wird die Annahme getroffen, dass dies ein notwendiges Kriterium für soziale Netzwerke ist.

### 2.2.2 Positive Gradkorrelation

Eine weitere Eigenschaft ist die Gradkorrelation. In sozialen Netzwerken gibt es Assortative Mixing [New02]. Das bedeutet, dass eine positive Korrelation zwischen dem Grad eines Knotens und dem Grad seiner Nachbarn existiert. Die Freunde einer Person mit vielen Bekanntschaften haben tendentiell auch viele Bekanntschaften, während die wenigen Freunde einer Person mit wenigen Bekannten auch nur schwach

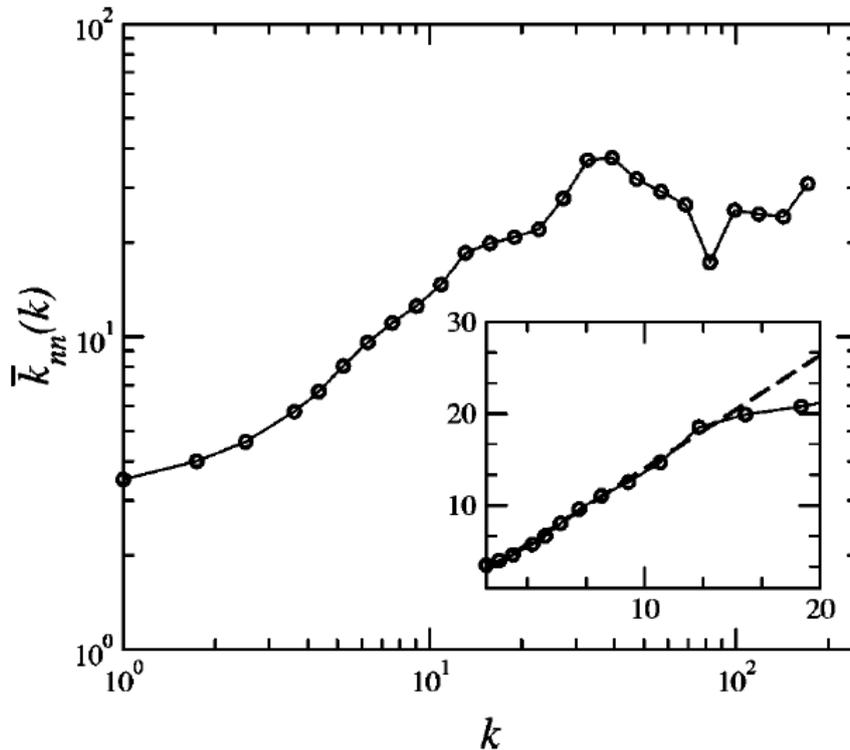


Abbildung 2: Durchschnittlicher Grad der direkten Nachbarn in Abhängigkeit zum Grad

vernetzt sind. Disassortative Mixing ist dagegen eine Eigenschaft von "unsozialen" Netzwerken. Ein Beispiel hierfür sind Server und Clients. Ein Server ist mit vielen Clients verbunden, diese sind aber nicht untereinander, sondern nur mit wenigen Servern verbunden, wir haben also eine negative Gradkorrelation. Im PGP Web of Trust kann man eine positive Korrelation messen. Vor allem für kleine Grade (die wegen ihrer Häufigkeit aber besonders relevant sind) gibt es einen positiven linearen Zusammenhang zwischen Grad und Grad der Nachbarn (siehe Abbildung 2). Wir verlangen also von den von uns erzeugten Netzwerken ebenfalls, dass es eine positive Korrelation gibt.

### 2.2.3 Community Structure

Die Gruppenstruktur in sozialen Netzwerken weist ganz bestimmte Muster auf. Sie ist allerdings nur schwer zu quantifizieren. Die Autoren verwenden hierfür den Algorithmus von Girvan und Newman [MG02]. Das ist ein Algorithmus, der einen Graph in seine Untergruppen aufteilt, in dem er nach und nach die Kanten entfernt, die besonders wichtig für den Zusammenhalt des Graphen sind. Das sind die Kanten mit der größten Betweenness Centrality, also die Kanten, die in den meisten kürzesten Pfaden enthalten sind. Das Resultat kann in einem Dendrogramm darge-

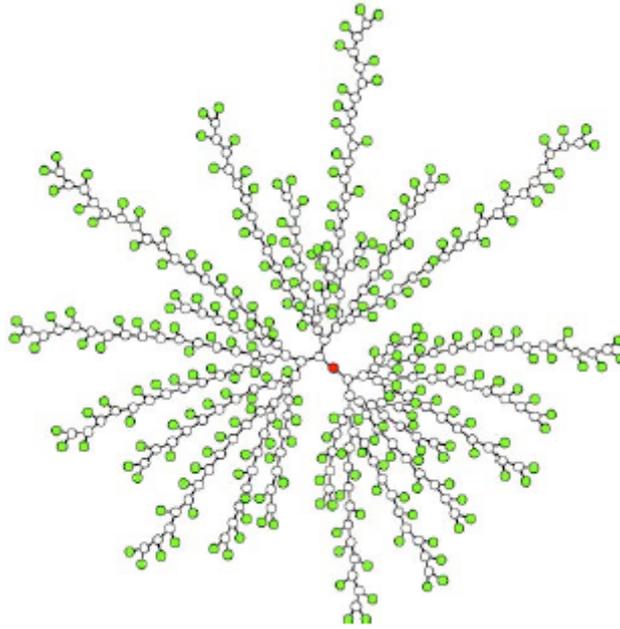


Abbildung 3: Beispiel eines Dendogramms. Der rote Knoten repräsentiert den ursprünglichen Graphen. Die weißen Knoten stellen die Gruppen dar, die durch die Aufteilung entstehen. Die grünen Knoten sind die tatsächlichen Knoten des Ursprungsgraphen

stellt werden. Das ist ein Baum, in dem jeder Knoten eine Gruppe repräsentiert und die Kinder des Knotens die beiden Gruppen darstellen, die sich durch die Aufteilung ergeben (siehe Abbildung 3).

Um das zu quantifizieren, kann man die Verteilung  $P$  der so resultierenden Gruppengrößen  $s$  messen. Beim PGP Web of Trust ist die Gruppengröße nach dem Potenzgesetz verteilt mit  $P(s) = s^{-1.8}$ . Wäre die Verteilung  $P(s) = s^{-2}$ , so würde jede Aufteilung der Gruppen 2 genau gleich große Gruppen ergeben. Man hätte also eine hierarchische Struktur von Gruppen und Untergruppen bis auf die Ebene der einzelnen Knoten. Wir erwarten also von unserem Modell, dass es ebenfalls hierarchische Strukturen mit einer ähnlichen Verteilung der Gruppengrößen erzeugt.

#### 2.2.4 Kritik

Auch wenn es im Paper niemals explizit behauptet wird, wird die Annahme vermittelt, dass diese 3 Eigenschaften nicht nur notwendig, sondern auch hinreichend für ein soziales Netzwerk sind. Meiner Meinung nach müsste man wesentlich mehr Analysen vornehmen, um ein soziales Netzwerk genau charakterisieren zu können. Außerdem ist vor allem die 3. Eigenschaft fragwürdig. Es muss auf jeden Fall eine bestimmte Gruppenstruktur in sozialen Netzwerken geben, vor allem da die Gruppenstruktur sowohl den Clustering Coefficient als auch die Korrelation des Grades

beeinflusst. Die Art und Weise, wie diese quantifiziert wird, macht wenig Sinn. Erstens ergibt jede Aufteilung in 2 ungefähr gleich große Gruppen eine Verteilung nach dem Potenzgesetz. Zweitens ist es mit dem Algorithmus auch möglich, bei komplett "unsozialen" Netzwerken (z.B. einem Pfad) ähnliche Ergebnisse zu erzielen. Auch wenn es wahrscheinlich keinen großen Einfluss auf das Ergebnis hat, sollte außerdem erwähnt werden, dass die benutzte Definition der Betweenness Centrality nicht ganz korrekt ist. Wenn zwischen 2 Knoten mehrere kürzeste Pfade existieren, erhöht sich die Betweenness Centrality normalerweise nur anteilig. Die absolute Anzahl der kürzesten Wege wird hingegen Stress Centrality genannt [UB05].

## 3 Das Social Space Modell

### 3.1 Raum

Dem Social Space Modell liegt die intuitive Annahme zu Grunde, dass Individuen nur dann eine Beziehung miteinander haben, wenn sie sich nahe stehen. Den Abstand zweier Personen kann man u.a. geografisch interpretieren, da die Wahrscheinlichkeit einer Bekanntschaft sicher stark vom Wohnort abhängt. Allerdings gibt es noch wesentlich mehr Faktoren, die diese Wahrscheinlichkeit beeinflussen, wie z.B. Hobbies, Beruf, Bildung oder Religion. Wir müssen also eine Möglichkeit finden, die Social Distance von zwei Individuen unter Berücksichtigung all dieser Faktoren zu messen und mit ihrer Hilfe die Verbindungswahrscheinlichkeit zu modellieren.

Um den Abstand zu berechnen, wird ein Social Space  $\mathcal{H}$  mit  $d_{\mathcal{H}}$  Unterräumen definiert. Jeder Unterraum repräsentiert eine Eigenschaft. Beispielsweise kann die geografische Position einer Person als Unterraum mit 2 Dimensionen (für Längen- und Breitengrad) dargestellt werden oder Religionszugehörigkeit als vieldimensionaler Unterraum, bei dem jede Religion eine Dimension repräsentiert. Wichtig ist, dass auf jedem Unterraum eine Metrik definiert ist, so dass man innerhalb eines Unterraums einen Abstand berechnen kann. In manchen Fällen kann dies eine simple (z.B. eine euklidische) Metrik sein. Bei den geografischen Standpunkten ist die Abstandberechnung zwar etwas komplizierter, aber immer noch problemlos möglich. Komplizierter wird es bei Abstrakteren Eigenschaften wie der Religion, da es hier keine offensichtliche Lösung des Problems gibt.

Jedes Individuum hat so eine eindeutige Position in  $\mathcal{H}$ . Seine Koordinaten in diesem Raum liefern eine vollständige Beschreibung all seiner Eigenschaften.

## 3.2 Abstraktion auf das Modell

Aus der Vorstellung von Individuen, die sich in in einem Raum  $\mathcal{H}$  befinden und über ihre Koordinaten beschrieben werden, müssen wir ein Modell konstruieren. Das geschieht folgendermaßen:

Wir platzieren  $N$  unverbundene Individuen gleichverteilt in einem social space  $\mathcal{H}$ . Jedes Individuum  $i$  ist eindeutig über den Vektor  $\vec{h}_i = (h_i^1, \dots, h_i^{d_{\mathcal{H}}})$  identifizierbar. Die Anzahl  $d_{\mathcal{H}}$  und die Art der Unterräume können wir frei wählen, solange alle Unterräume eine Metrik besitzen. Auch Modelle mit  $d_{\mathcal{H}} = 1$  sind denkbar. Da die Individuen gleichverteilt platziert werden, sind die verschiedenen Unterräume unkorreliert. Daraus folgt, dass die Wahrscheinlichkeit  $p(\vec{h})$  für die Existenz eines Individuums  $\vec{h}$  folgendermaßen berechnet werden kann:  $p(\vec{h}) = \prod_{n=1}^{d_{\mathcal{H}}} p_n(h^n)$ .

## 3.3 Verbindungswahrscheinlichkeit

Wir berechnen nun paarweise für alle Individuen die Wahrscheinlichkeit, dass sie sich verbinden. Diese Wahrscheinlichkeit berechnen wir über die Distanzen in den Unterräumen. Für je 2 Individuen  $\vec{h}_i$  und  $\vec{h}_j$  ist die Distanz  $d_n$  in jedem Unterraum  $n$  gemäß der jeweiligen Metrik  $d_n(h_i^n, h_j^n)$ . Anhand dieser Distanz berechnen wir zunächst die Wahrscheinlichkeit, dass 2 Individuen innerhalb eines Unterraums verbunden sind. Diese Wahrscheinlichkeit ist:

$$r_n(h_i^n, h_j^n) = \frac{1}{1 + [\frac{1}{b_n} d_n(h_i^n, h_j^n)]^{\alpha_n}}$$

Die beiden vom Unterraum abhängigen Parameter  $\alpha_n$  und  $b_n$  haben folgende Funktion:

- Length Scale  $b_n$ : Durch  $b_n$  kann jeder Unterraum skaliert werden. Umso größer  $b_n$  ist, desto kleiner wird die Wahrscheinlichkeit einer Verbindung innerhalb des entsprechenden Unterraums. Die length scale hat somit einen direkten Einfluss auf  $\langle k \rangle$ .
- Homophily  $\alpha_n$ : Die Homophily  $\alpha_n > 1$  beschreibt die Tendenz eines Individuums, sich bevorzugt mit nahen Individuen zu verbinden. Da die Homophily die Distanzen exponiert, werden kleine Distanzen weniger stark vergrößert als große Distanzen.

Man kann sehr schön sehen, dass die Wahrscheinlichkeit mit steigender Distanz abnimmt. Wenn unser social space nur einen Unterraum besitzt, ist diese Formel so wie sie ist ausreichend. Die Beispiele im nächsten Kapitel zeigen, dass man auch mit nur

einem Unterraum realistische Ergebnisse erzeugen kann. Bei mehreren Unterräumen muss lediglich die Summe über die Wahrscheinlichkeit aller Unterräume berechnet werden.

$$r(\vec{h}_i, \vec{h}_j) = \sum_{n=1}^{d_H} \omega_n r_n(h_i^n, h_j^n)$$

Der Parameter  $\omega_n$  gewichtet die verschiedenen Unterräume, da nicht alle Unterräume die Wahrscheinlichkeit gleich stark beeinflussen. Der geografische Standpunkt hat beispielsweise einen wesentlich größeren Einfluss als die Religionszugehörigkeit.

### 3.4 Kritik am Aufbau des Social Space Modells

Alle Individuen gemäß ihrer Eigenschaften in einem Social Space zu platzieren und anhand ihrer Koordinaten die Verbindungswahrscheinlichkeit zu berechnen klingt zunächst naheliegend. Was aber auf jeden Fall hinterfragt werden muss, ist die Tatsache, dass die Summe von verschiedenen Wahrscheinlichkeiten auch Werte  $> 1$  ergeben kann. Wenn man davon ausgeht, dass eine Menge von Unterräumen alleine schon so dominant ist, dass die übrigen Unterräume keinen Einfluss mehr auf das Ergebnis haben, ist dies aber erklärbar. Zum Beispiel werden sich 2 Schüler der selben Klasse auf jeden Fall unabhängig von ihrer Religion kennen. Außerdem kann durch sinnvolle Wahl der Parameter verhindert werden, dass sich unsinnige Wahrscheinlichkeiten ergeben.

## 4 Analyse der erzeugten Netzwerke

Wie oben schon erwähnt, lassen sich auch mit nur einem Unterraum schon viele soziale Netzwerke erzeugen. Abbildung 4 zeigt 3 sehr unterschiedliche Netzwerke, die mit nur einem Unterraum, allein durch die Variation der Homophily und des Length Scale entstanden sind.

Im folgenden werden unterschiedliche Netzwerke mit variierenden Parametern erzeugt. Es wird untersucht, welchen Einfluss diese Parameter auf die für soziale Netzwerke wichtigen Eigenschaften (Clustering Coefficient, Korrelation des Knotengrads und Community Structure) haben und ob es möglich ist, diese Charakteristiken nachzubilden.

### 4.1 Clustering Coefficient

Anhand des PGP Web of Trust haben wir festgestellt, dass der Clustering Coefficient eines sozialen Netzwerks 2 wichtige Eigenschaften erfüllen muss.

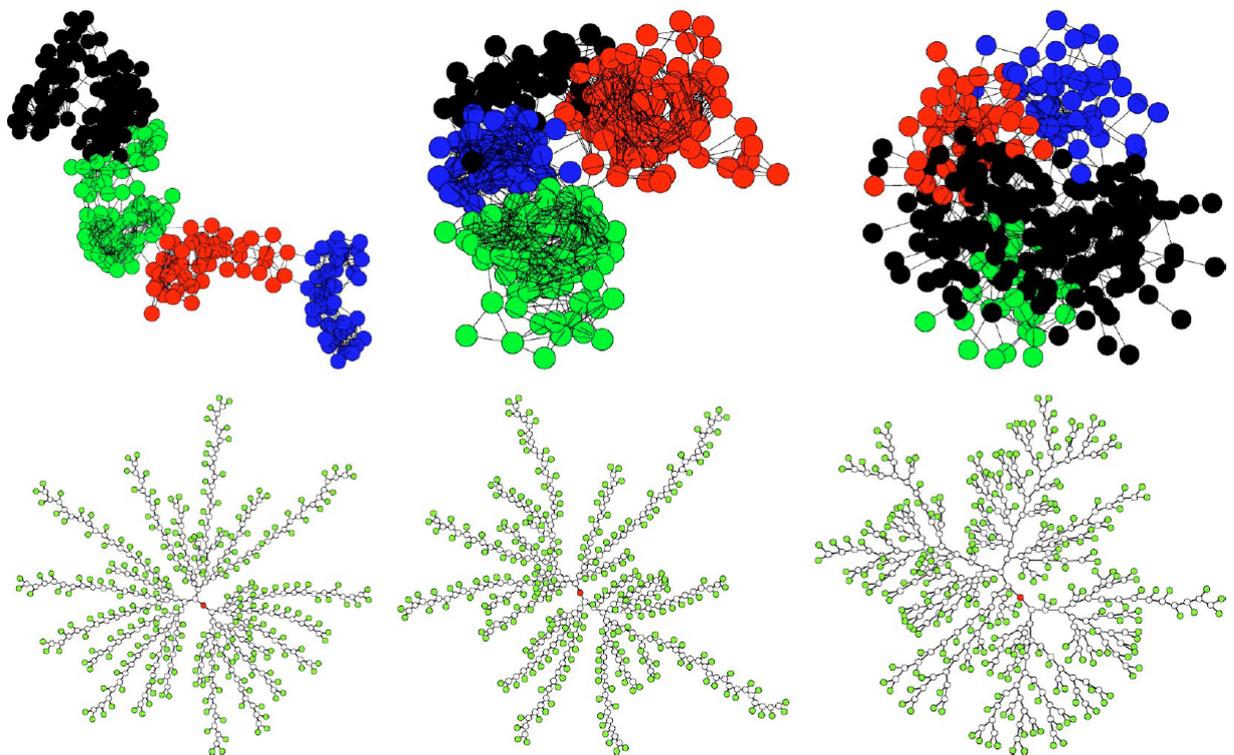


Abbildung 4: Oben: Beispielnetzwerke mit einem Unterraum und variierender Homophily bzw. Length Scale. Beim rechten Netzwerk ist  $\alpha$  ungefähr 1, nach links steigt  $\alpha$  an. Die Farben unterscheiden die ersten 4 Gruppen nach Aufteilung durch Girvan und Newman. Unten sieht man die zugehörigen Dendogramme

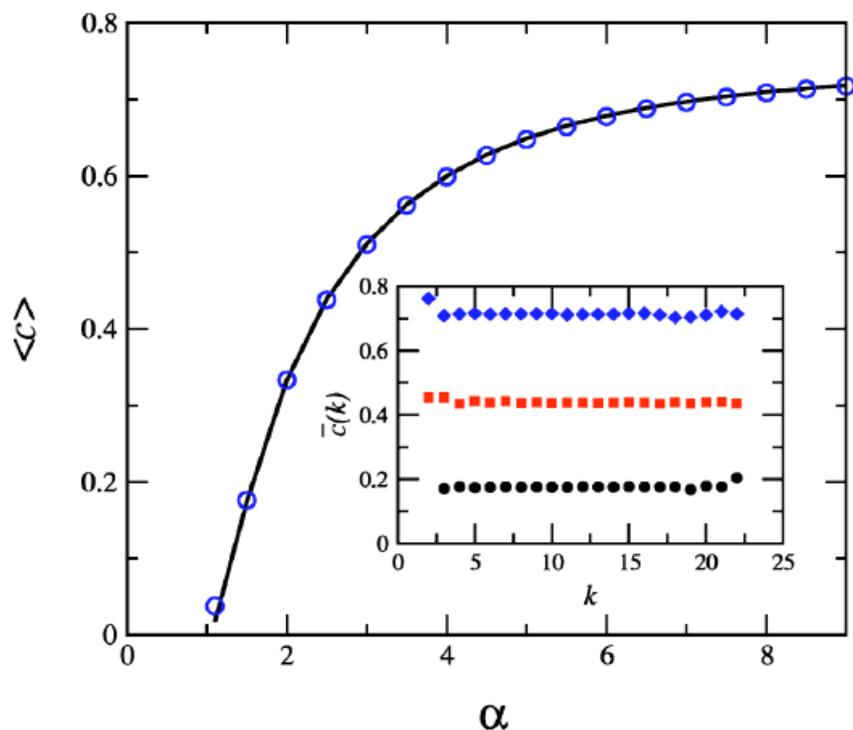


Abbildung 5: Großes Diagramm:  $\langle c \rangle$  in Abhängigkeit zu  $\alpha$ , kleines Diagramm:  $\langle c \rangle$  in Abhängigkeit zum Knotengrad für 3 verschiedene Werte von  $\alpha$

- $\langle c \rangle$  besitzt einen relevanten Wert, der höher ist als bei zufälliger Verbindung der Knoten.
- $\langle c \rangle$  ist für relevante Knotengrade unabhängig von diesem.

Das große Diagramm in Abbildung 5 zeigt, dass  $\langle c \rangle$  sehr hohe Werte haben kann, und somit auf jeden Fall relevant ist. Der Clustering Coefficient ist sehr stark von  $\alpha$  abhängig. Dies ist naheliegend, da durch ein großes  $\alpha$  Verbindungen zu nahen Knoten bevorzugt werden, und somit auch transitive Beziehungen. Das kleine Diagramm in Abbildung 5 zeigt  $\langle c \rangle$  in Abhängigkeit zum Knotengrad für 3 verschiedene Werte von  $\alpha$ . Auch hier sieht man sehr deutlich, dass unsere Bedingungen für soziale Netzwerke erfüllt werden, denn für alle  $\alpha$  ist der Clustering Coefficient nahezu unabhängig vom Knotengrad.

## 4.2 Korrelation des Knotengrades

Eine weitere Eigenschaft sozialer Netzwerke, die beim PGP Web of Trust festgestellt werden konnte, ist Assortative Mixing des Knotengrads, also eine positive Korrelation zwischen dem Grad eines Knotens und dem durchschnittlichen Grad seiner Nachbarn. Abbildung 6 zeigt diese Korrelation für 3 verschiedene Netzwerke mit

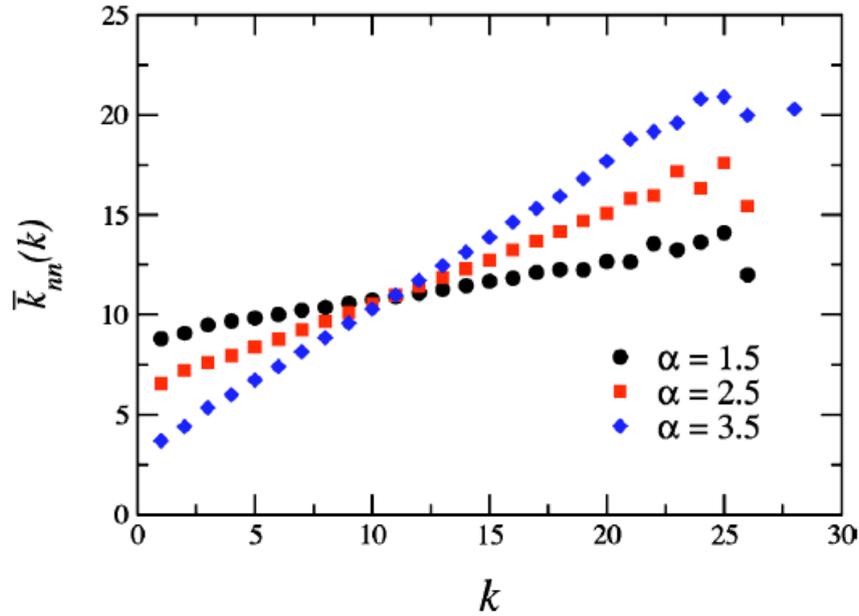


Abbildung 6: Durchschnittlicher Grad aller Nachbarn eines Knotens in Abhängigkeit zu seinem Grad

unterschiedlicher Homophily. Wie auch schon beim PGP Web of Trust existiert für relevante Knotengrade ein beinahe linearer Zusammenhang, dessen Steigung von  $\alpha$  abhängig ist. Ein höherer Clustering Coefficient bewirkt eine höhere Korrelation.

### 4.3 Community Structure

Ein weiteres Ziel war es, Netzwerke zu erzeugen, die eine hierarchische Gruppenstruktur aufweisen. Diese wird anhand der Verteilung der Gruppengrößen bei einer Aufspaltung des Netzwerkes durch den Algorithmus von Girvan und Newman gemessen. Abbildung 7 zeigt die Verteilung der Gruppengrößen für 3 unterschiedliche Netzwerke mit unterschiedlichem  $\alpha$ . Umso geringer die Homophily ist, desto besser lässt sich die hierarchische Struktur (gekennzeichnet durch  $P(s) = s^{-2}$ ) erkennen. Diese sieht man auch an dem rechten Zerlegungs dendogrammen aus Abbildung 4, das sehr regelmäßige Aufteilungsstrukturen aufweist. Wenn  $\alpha$  jedoch höher ist, weicht die Struktur deutlich davon ab. Im linken Dendogramm sieht man gar keine Hierarchien mehr, sondern eher eine Art Verkettung. Für größere Netzwerke soll laut den Autoren aber auch hier eine hierarchische Struktur erkennbar sein.

### 4.4 Kritik an der Netzwerkanalyse

Die Autoren kommen zu dem Ergebnis, dass sich alle wichtigen Eigenschaften sozialer Netzwerke durch die Parameter im Social Space Modell beeinflussen lassen

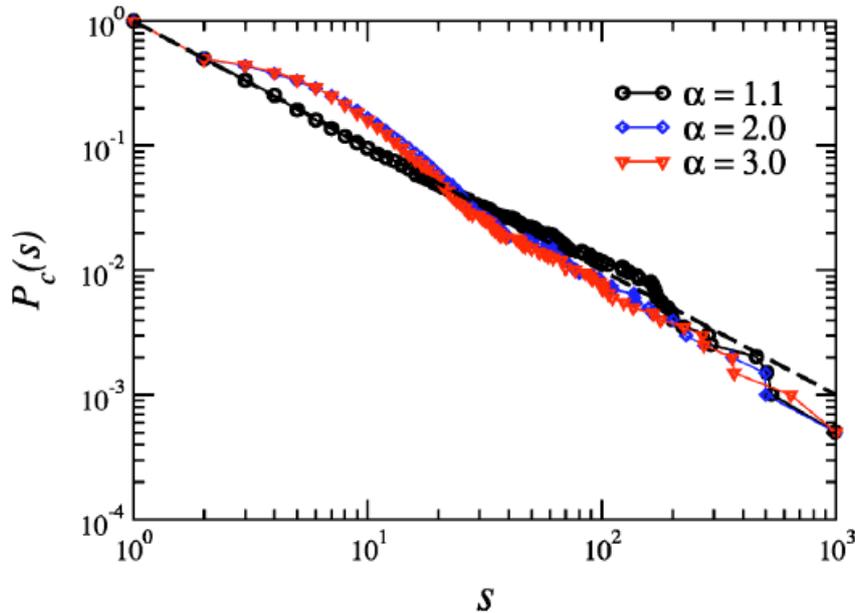


Abbildung 7: Verteilung der Gruppengrößen, die sich durch Aufteilung mit dem Algorithmus von Girvan und Newman ergibt für unterschiedliche  $\alpha$

und werten dies als Beweis dafür, dass man mit dem Social Space Modell soziale Netzwerke modellieren kann. Allerdings halten sie sich selber nicht ganz an die von ihnen gemachten Vorgaben aus dem PGP Web of Trust. Das wird vor allem bei der Community Structure deutlich. Sie beweisen lediglich, dass sie in der Lage sind, gewisse Hierarchien zu erzeugen, indem sie zeigen, dass die Gruppenverteilung  $P(s) = s^{-2}$  ist. Diese lässt sich zwar noch durch  $\alpha$  beeinflussen, aber einen direkten Einfluss auf den Exponenten des Potenzgesetzes haben sie nicht. Nach ihren eigenen Vorraussetzungen müssten sie aber in der Lage sein, eine Verteilung  $P(s) = s^{-1.8}$  zu erzeugen, das bleiben sie uns schuldig. Nachdem sie in der Einführung alle wichtigen Eigenschaften sozialer Netzwerke aus einem einzigen realen Netzwerk abgeleitet haben, wäre es außerdem sehr schön gewesen, wenn sie ein Netzwerk erzeugt hätten, was in allen Eigenschaften dem PGP Web of Trust entsprochen hätte. Stattdessen bleibt bei mir hingegen der Eindruck, dass das garnicht möglich ist, denn ein hohes  $\alpha$ , wie für die Korrelation des Grads gefordert, hat negativen Einfluss auf die hierarchischen Gruppenstrukturen.

#### 4.5 Verwendung im Attract & Introduce Modell

Das Attract & Introduce Modell hat grundsätzlich das selbe Ziel, man möchte möglichst realistische soziale Netzwerke erzeugen. Während beim Social Space Modell das PGP Web of Trust als Vergleich benutzt wurde, stützen sich die Autoren

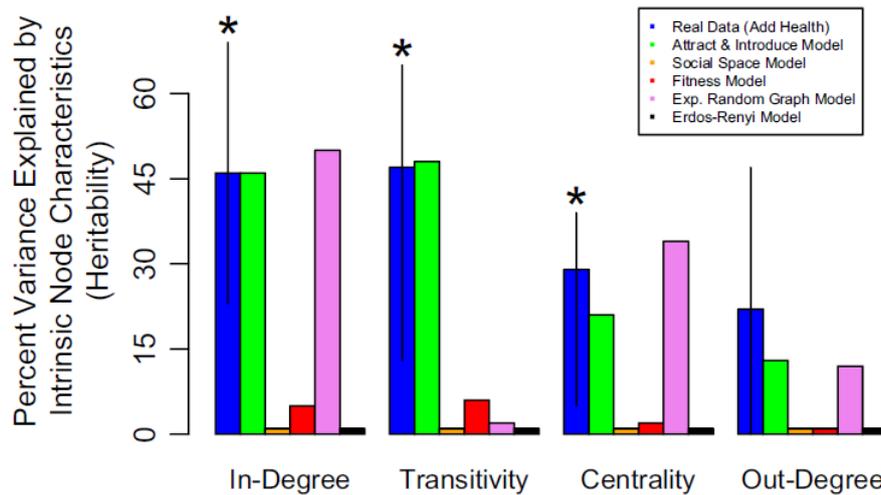


Abbildung 8: Vererbbarkeit verschiedener Netzwerkeigenschaften in verschiedenen Modellen

des Attract & Introduce Modells auf die Add Health Studie. Diese Studie enthält ein großes soziales Netzwerk. Anhand der Zwillinge in dem Netzwerk konnte gezeigt werden, dass bestimmte Netzwerkeigenschaften wie In-Degree, Betweenness Centrality oder Transitivität vererbbar sind. Die Autoren gehen deshalb davon aus, dass jedes Modell, das realistische soziale Netzwerke erzeugt, ebenfalls Knotenbezogene Netzwerkeigenschaften vererben muss. Sie haben die "Mirror Network" Methode eingeführt, um verschiedene Netzwerkmodelle darauf zu testen, ob mit ihnen Vererbung möglich ist. Das bei diesem Test das Attract & Introduce Modell am nächsten an die realen Werte der Add Health Studie rankommt, ist wenig überraschend, da es genau dafür designed wurde. Das Social Space Modell schneidet in diesem Test besonders schlecht ab, es ist nahezu keine Vererbung messbar (siehe Abbildung 8). Das ist auch nicht besonders verwunderlich, denn alles was ein Knoten in dem Modell vererben kann, sind seine Koordinaten im Social Space. Da diese gleichverteilt sind, können die wichtigen Netzwerkeigenschaften nicht vererbt werden. Lediglich bei Knoten am Rand des Social Space kann es zu minimalen Vererbungseffekten kommen, denn die Randknoten haben modellbedingt einen niedrigeren Grad, diese Eigenschaft bleibt auch bei Übertragung in ein neues Netzwerk bestehen.

## 5 Fazit

Die intuitive Annahme, dass Personen sich bevorzugt zu ihnen nahe stehenden Personen verbinden, überzeugt schnell und macht das Social Space Modell auf den ersten Blick sehr sympatisch. Leider kann das Paper selbst an einigen Stellen nicht

ganz überzeugen. Vor allem die Charakterisierung sozialer Netzwerke und die Auswertung der erzeugten Netzwerke sind in sich nicht immer schlüssig. Dass das Social Space Modell in der "Mirror Network" Methode sehr schlecht abschneidet, ist wenig überraschend und auch kein großer Kritikpunkt, denn diese Methode wurde für ganz andere Netzwerke entworfen. Es ist generell schwierig das perfekte Modell zu finden, weil bei der Vielfalt an Analysemethoden kein Modell immer am besten abschneiden kann. Vielmehr sollte man meiner Meinung nach die Mechanismen, die in den Modellen arbeiten versuchen zu kombinieren. So könnte man beispielsweise zur Berechnung der Verbindungswahrscheinlichkeit im Social Space Modell noch Attract und Introduce Werte hinzuziehen.

## Literatur

- [Add] Add health. <http://www.cpc.unc.edu/projects/addhealth>.
- [Fei02] Patrick Feisthammel. Das web of trust. <http://www.rubin.ch/pgp/weboftrust.de.html>, 2002.
- [JHF09] Nicholas A. Christakis James H. Fowler, Christopher T. Dawes. Model of genetic variation in human social networks. *Proceedings of the National Academy of Sciences*, 106(6):1720-1724, 2009.
- [MBn04] Albert Díaz Guilera Alex Arenas Marían Boguñá, Romualdo Pastor-Satorras. Models of social networks based on social distance attachment. *PHYSICAL REVIEW*, E 70, 056122, 2004.
- [MG02] M. E. J. Newman M. Girvan. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821-7826, 2002.
- [New02] M. E. J. Newman. Assortative mixing in networks. *PHYSICAL REVIEW*, Lett. 89, 208701, 2002.
- [UB05] Thomas Erlebach Ulrik Brandes. *Network Analysis Methodological Foundations*. Springer Verlag, 2005.