

Assignment Social Network Analysis

Meta Infos

Student	Jonathan Baß
Titel	Analyse der sozialen Strukturen einer Krankenhausabteilung mit Fokus auf das epidemiologische Potential der Rollen
Kurskontext	Social Network Analysis
Datum	18.11.2021
Dozent	Philipp M. Mendoza, M.Sc.

Executive Summary

- Einleitung
 - Epidemiologie durch Corona-Pandemie aktueller denn je
 - Bezug nicht nur auf medizinische Einrichtungen, sondern auch auf das allgemeine Leben
 - Personen mit vielen Kontakten können als Superspreader fungieren und Krankheiten schnell verbreiten
- Forschungsfrage
 - Bestimmung von Berufsgruppen mit besonders hohem Risiko der Krankheitsverbreitung
 - Einfluss von räumlich gegebenen Trennungen der Personen
 - Risiko der Rollen bestimmen und soziale Struktur auf Problemstellen überprüfen
- Datensatz
 - Daten wurden in geriatrischer Abteilung eines französischen Krankenhauses bei Experiment erhoben
 - Dabei wurden Begegnungen zwischen Personen auf der Station, sowie deren Rollen erfasst
 - Datensatz umfasst 46 Mitarbeiter des Gesundheitswesens, 29 Patienten und insgesamt ca. 14.000 Kontakte
- Strategie
 - Abstrakte Betrachtung der Community-Struktur durch Anwendung von Annealing-Algorithmus “Netcarto”
 - Visualisierung der Ergebnisse mit Fokus auf die Schlüsselwerte “Within-community Degree” und “Participation Coefficient”
 - Darstellung des Netzwerkes in Kombination mit den Daten aus dem Algorithmus
- Ergebnisse
 - Gesundheitspersonal besonders anfällig ein Superspreader zu sein
 - Verwaltungspersonal darf in den Planungen nicht vernachlässigt werden
 - Aktive Bildung von Communities, durch Aufteilung der Krankenpfleger/innen und Ärzt/innen senkt das Risiko einer großen Verbreitung

Einleitung und Fragestellung

Die Analyse und Nachverfolgung von Kontakten ist in der Medizin und Forschung schon lange ein wichtiges und fundamentales Forschungsgebiet (Epidemiologie), wenn es um die Verbreitung von Krankheiten in medizinischen Bereichen wie Krankenhäusern oder Laboren geht. Durch die SARS-CoV-2-Pandemie wurde der Fokus dieses Forschungsbereiches jedoch auch auf das alltägliche Leben erweitert, denn bei engen und/oder langen Kontakten können Keime und Erreger, wie der Corona Virus, zwischen den Personen ausgetauscht werden. Daher ist das Thema und dessen Analyse wichtiger und aktueller denn je. Die folgende Analyse geht von einer Situation ohne akute Pandemie aus. Am Ende werden die Ergebnisse auf die aktuelle Situation übertragen.

Es gibt viele Erreger, die durch Kontakt zwischen Personen übertragen werden und sich Krankheiten auf diesem Weg verbreiten können. Dies stellt jedoch im Alltag normalerweise kein großes Problem dar, da die meisten Erreger harmlos sind und das menschliche Immunsystem den Ausbruch der Krankheit verhindern kann. Dies ist in Krankenhäusern aber nicht immer der Fall. Durch die hohe Konzentration an angewendeten Medikamenten wie Antibiotika können sich schnell multiresistente Keime bilden, welche auch Krankenhauskeime genannt werden, und nur sehr schwer behandelt werden können. Zudem haben Patienten meist ein geschwächtes Immunsystem. Aus diesem Grund kann eine Infektion mit solchen Erregern tödlich verlaufen. Daher wird in solchen Umgebungen besonders auf die Hygiene geachtet, um eine Verbreitung möglichst zu vermeiden. Eine detaillierte Beschreibung und Quantifizierung der Kontakte in Krankenhäusern kann deshalb wichtige Informationen für die Epidemiologie von Krankenhausinfektionen sowie für die Konzeption und Validierung von Kontrollmaßnahmen liefern. Wie diese vorgenommen werden und welche Schlüsse aus der Analyse gezogen werden können, wird in dem folgenden Beitrag erläutert.

Thema

In der geriatrischen Abteilung eines Krankenhauses in Lyon, Frankreich, wurde 2010 ein Experiment gemacht. Bei diesem wurden alle sich auf der Station befindlichen Personen mit tragbare RFID-Sensoren (Radio Frequency Identification) ausgerüstet, welche alle Interaktionen im Nahbereich von etwa 1,5m und einer zeitlichen Auflösung von 20 Sekunden über einem Zeitraum von vier Tagen und vier Nächten gemessen und aufgezeichnet haben. Die Studie umfasste 46 Mitarbeiter des Gesundheitswesens und 29 Patienten. Dabei wurden insgesamt über 14.000 Kontakte erfasst. Da gerade eine geriatrische Abteilung sehr pflegeintensive Patienten besitzt, welche zudem eine hohe Anfälligkeit haben, eignet sich der Datensatz sehr gut für die angestrebte Analyse.

Daten

Die Daten liegen als ein `igraph graph`-Objekt vor, welches die `graph`-Attribute “name” und “Citation”, als `vertex`-Attribut “Status” und als `edge`-Attribut “Time” besitzt. “Status” beschreibt dabei die Rolle der Person. Dabei wird in Verwaltungspersonal (*ADM*), Ärzte (*MED*), medizinisches Personal wie Krankenschwestern und -Pfleger oder Hilfspersonal (*NUR*) und Patienten/Patientinnen (*PAT*) unterschieden. “Time” ist der Zeitstempel der Sekunde, an dem das 20-Sekunden Intervall ausgelaufen und nicht erneuert wurde. Für die Analyse werden die Zeitstempel nicht weiterverwendet, da von einer direkten Infizierung bei auch schon kurzen Begegnungen ausgegangen wird.

Forschungsfrage

Die Übertragung von Krankheiten kann unter anderem durch persönlichen Kontakt in Form einer Tröpfchen- oder Schmierinfektion stattfinden. Personen, die viele Kontakte mit einer Vielzahl an Personen haben, können dabei als Superspreader fungieren und Krankheiten schnell verteilen. **In der folgenden Ausarbeitung beschäftige ich mich mit der Frage, ob gewisse Rollen in Krankenhäusern ein höheres Risiko haben, solch ein Superspreader zu sein, und ob die durch organisatorische und räumlich gegebene Bildung von Gemeinschaften einen Einfluss auf eine mögliche Verbreitung hat.**

Relevanz der Forschungsfrage

Die Kontaktnachverfolgung ist durch die aktuelle Corona-Pandemie ein wichtigeres Thema denn je. Es gilt geschwächte und anfällige Personen zu schützen. Gerade in Krankenhäusern kann eine unkontrollierte Ausbreitung des SARS-CoV-2-Virus durch einen Superspreader verheerende Auswirkungen haben. Die Analyse und Beantwortung der Forschungsfrage kann bei der Entwicklung eines Schutzkonzeptes helfen und das allgemeine Risiko einer Infektion während eines Krankenhausaufenthaltes mit gefährlichen Erregern verringern. Dabei sind gerade Daten aus dieser Abteilung von großem Nutzen, da hier ausschließlich ältere und geschwächte Patienten behandelt werden.

Analysestrategie

Um die Forschungsfrage korrekt und vollständig beantworten zu können, reicht es nicht aus, die reinen Begegnungen zwischen den Personen zu zählen. Sollte zum Beispiel ein infizierter Patient viel Kontakt mit einem anderen infizierten Patienten haben, würde dadurch keine Übertragung und damit auch keine weitere Ausbreitung der Krankheit erfolgen. Aus diesem Grund muss das Verhältnis der Personen zu anderen Personen abstrahiert werden und die Gemeinschaftsstrukturen genauer analysiert werden. Dafür können zwei Werte für jede Node berechnet werden: Der “Within-community Degree”-Wert (deutsch: Innergemeinschaftlicher Grad) beschreibt die Vernetzung innerhalb der eigenen Community. Dazu wird der “Participation Coefficient” (deutsch: Teilnahmekoeffizient) betrachtet. Dieser beschreibt die Vernetzung einer Person über die Grenzen der eigenen Community hinaus. Für jede Node können beide Werte berechnet und auf einem Diagramm als Punkte in ein Verhältnis gebracht werden. Sollte zum Beispiel eine Person eine sehr gute Vernetzung innerhalb der eigenen Community aber auch zu anderen Communities haben, ist das Risiko, bei einer Infektion als Superspreader zu fungieren, stark erhöht. Für die Berechnung dieser Schlüsselwerte wird der Algorithmus “Netcarto” verwendet. Netcarto bietet schnelle Netzwerkmodularitäts- und Rollenberechnungen durch simuliertes Annealing. Die Ergebnisse werden dann in dem zweiten Teil als Edge-Data dem Netzwerk zugeführt und visualisiert.

Charakterisierung des Netzwerks

Der Datensatz bildet die Interaktionen innerhalb der Station über einen Zeitraum von 96 Stunden ab. Die 75 Nodes (Vertices) repräsentieren die Teilnehmer des Experiments und sind alle von dem gleichen Typ. Die Node-Daten beinhalten jeweils den Status bzw. die Rolle der Person, welche im Punkt “Daten” bereits beschrieben sind. Die Edges (Links) repräsentieren die Kontakte der Personen untereinander. Je öfter zwei Personen Kontakt miteinander hatten, desto mehr Edges gibt es zwischen ihnen. Da die Links nur den reinen Kontakt symbolisieren sind sie nicht gerichtet (directed) und zeigen nur, dass ein engerer Kontakt stattgefunden hat. Aus diesem Grund findet auch keine Gewichtung (weighted) der Verbindungen statt. Zudem können für die reine Analyse alle Schleifen entfernt werden, da bereits ein Kontakt für eine potenzielle Übertragung ausreicht. Nicht benötigte Daten werden entfernt, um die Berechnungen zu beschleunigen. Zusammenfassend lässt sich sagen, dass es sich um ein One-Node-Netzwerk handelt, welches weder eine Richtung noch eine Gewichtung in den Nodes besitzt.

Umsetzung

Datenmanipulationen

Erklärung der Schritte

Die folgende Auflistung beschreibt die einzelnen Schritte, welche für die Auswertung verwendet wurden. Diese können chronologisch ausgeführt werden, um zu dem gleichen Ergebnis zu gelangen.

1. Umgebung Vorbereiten

Zu Beginn müssen alle benötigten Pakete installiert werden. Darunter befinden sich unter anderem “igraph-data”, aus welchem die Daten geladen werden und standart-Pakete wie “ggraph” oder „GGally“, welche für die Visualisierungen verwendet werden. Zudem benötigt man den oben erwähnten Annealing-Algorithmus “rnetcarto”. Für die Installationsverwaltung der Pakete wird die folgende Funktion verwendet. Diese Prüft, ob alle in der Liste “uP” definierten Paktete installiert sind und führt die Installations-Anweisung nur noch für die fehlenden Pakete durch. Zudem werden die Pakete auch gleich als Bibliotheken deklariert und damit in der Umgebung auch instanziiert.

```
dP<- function(...){
  x <- list(...)
  for(i in x) {
    if (!is.element(i, installed.packages()[,1])) {
      install.packages(p, dep = TRUE, repos = "http://cran.wu.ac.at")
    }
    library(i, character.only = TRUE)
  }
}

dP('igraph', 'igraphdata', 'tidygraph', 'tidyverse', 'ggraph', 'GGally', 'RColorBrewer', 'rnetcarto')
```

Damit keine Konflikte mit alten Daten auftreten und die Berechnungen möglichst performant laufen können, werden alle alten Daten aus dem Environment gelöscht.

```
rm(list = ls())
```

2. Daten für den Algorithmus vorbereiten

Jetzt instanziiieren wir die Daten mit dem folgenden Befehl. Es existiert nun ein igraph graph-Objekt “rfid”, welches die benötigten Daten beinhaltet. Dieses Objekt bildet die Grundlage für die folgenden Berechnungen.

```
data("rfid")
```

Folgend werden alle Schleifen entfernt, um Fehler in den Darstellungen zu vermeiden. Die gesäuberten Daten werden in einer neuen Variable abgelegt.

```
df <- as.undirected(simplify(rfid))
```

Damit der Netcarto-Algorithmus die Daten verarbeiten kann, müssen diese als Adjazenzmatrix vorliegen. Der folgende Befehl konvertiert das Dateframe in dieses Format und speichert die neue Matrix als Variation der Variable df ab.

```
df.mat=as_adjacency_matrix(df, sparse = F)
```

3. Simulierter annealing Algorithmus

Nun kann der Netcarto-Algorithmus angewandt werden. Dieser simuliert ein sogenanntes “annealing” oder auf Deutsch “abkühlen”, was sich auf die Art und Weise bezieht, wie die einzelnen Communities berechnet werden. Eine gute Analogie ist das Abkühlen von Metall, nachdem es zum Glühen gebracht wurde. Zu diesem Zeitpunkt bewegen sich die einzelnen Atome noch sehr schnell, da sie ein hohes Energiepotential besitzen. Im Verlauf des Abkühlprozesses ordnen sich die Atome in einer kristallinen Form an und es wird ein nahezu optimaler Zustand erreicht. Energiereichere Atome benötigen für diesen Prozess länger. Zieht man nun die Verbindungen zu dem Algorithmus, so wird das Netzwerk in einen ähnlichen Zustand gebracht und es können einzelne Module mit einer Wahrscheinlichkeit von bis zu 90% berechnet werden. Dabei ist die Anzahl der Kontakte das Energiepotential.

```
rnc=netcarto(df.mat)
head(rnc[[1]], 5)
```

##	name	module	connectivity	participation	role
## 59	59	0	-1.8424238	0.6562500	Connector
## 75	75	0	-1.4591996	0.6115702	Peripheral
## 60	60	0	-1.2675875	0.6428571	Connector
## 56	56	0	-1.0759755	0.6222222	Connector
## 54	54	0	-0.8843634	0.6093750	Peripheral

4. Daten für den ersten Plot vorbereiten

Es liegt nun ein Dataframe mit den zu Beginn definierten Werten vor, welches auch das Modul beinhaltet. Leider ist jedoch beim Erstellen der Adjacency-Matrix die Rolle verloren gegangen, da es sich bei der Spalte um nicht-nummerische Werte handelt. Daher müssen die beiden Dataframes vereint werden. Dafür wird dem neu erstellten Dataframe eine Spalte mit dem Index hinzugefügt, um im nächsten Schritt einen merge durchführen zu können. Da diese Spalte bei dem ursprünglichen DF “name” heißt, wurde hier die gleiche Bezeichnung verwendet.

```
df2 <- df %>% as_tbl_graph() %>% activate(nodes) %>% mutate(name = row_number())
```

Durch das Hinzufügen der Index-Spalte kann nun ein Merge durchgeführt werden. Das Dataframe “df-Plot” beinhaltet jetzt alle benötigten Werte, welche für die Auswertungen relevant sind und kann in einem Diagramm visualisiert werden.

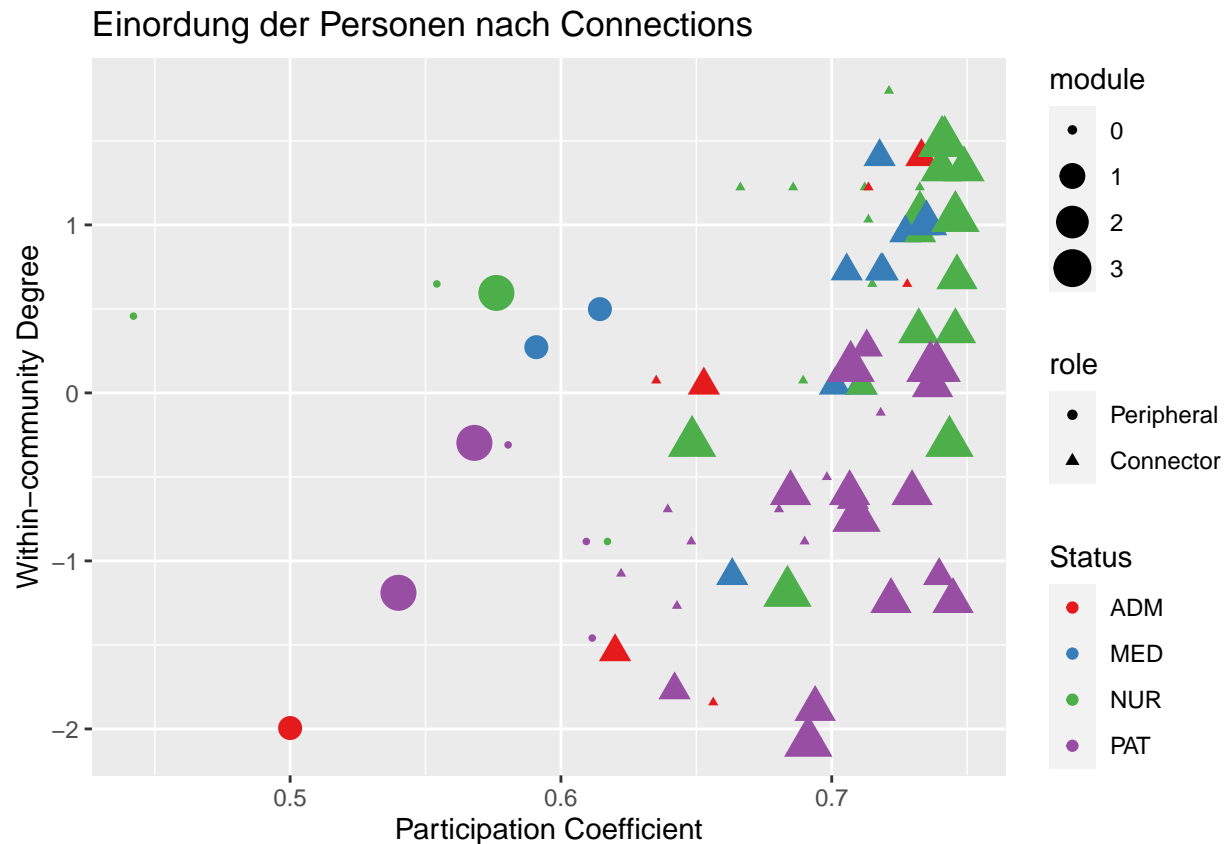
```
dfPlot <- merge(x = df2 , y = rnc[[1]], by = "name", all = TRUE)
head(dfPlot, 10)
```

##	name	Status	module	connectivity	participation	role
## 1	1	ADM	1	1.40546577	0.7331363	Connector
## 2	2	NUR	2	0.36706517	0.7321429	Connector
## 3	3	NUR	3	-0.29748206	0.7434842	Connector
## 4	4	NUR	2	0.36706517	0.7456747	Connector
## 5	5	NUR	3	1.04118721	0.7325620	Connector
## 6	6	NUR	3	1.48741029	0.7407407	Connector
## 7	7	NUR	2	1.00942923	0.7325331	Connector
## 8	8	NUR	3	0.59496412	0.5761773	Peripheral
## 9	9	MED	1	0.04533761	0.7011719	Connector
## 10	10	NUR	0	1.22336938	0.6662701	Connector

5. Plot #1

Der nächste Schritt ist das Plotten mit ggplot. Als y-Wert wird der “Within-community Degree” genommen, also der Wert, welcher die Vernetzung innerhalb der eigenen Community darstellt. Auf der x-Achse befindet sich “Participation Coefficient” bzw. die Vernetzung zu anderen Communities. Der Algorithmus unterscheidet dabei in verschiedene Rollen. Menschen, die weder eine hohe Vernetzung innerhalb noch außerhalb ihrer Community haben werden als “*Peripherals*” bezeichnet. “*Connectors*” spielen keine große Rolle innerhalb ihrer Community. Sie sind jedoch sehr gut mit anderen Vernetzt. Des Weiteren gibt es noch zwei andere Rollen, welche bei diesem Beispiel jedoch nicht auftreten. “*Provincial Hubs*” sind zum Beispiel nur in Ihrer Community gut vernetzt und “*Connector Hubs*” vereinen “*Connectors*” und “*Provincial Hubs*”.

```
ggplot(dfPlot, aes(y = connectivity, x = participation)) +  
  ggtitle("Einordnung der Personen nach Connections") +  
  xlab("Participation Coefficient") +  
  ylab("Within-community Degree") +  
  geom_point(aes(color=Status, shape=role, size=module)) +  
  scale_color_brewer(palette = "Set1")
```



6. Vorbereitung Kombination

Die vom Algorithmus berechneten Daten können nun mit dem Netzwerk kombiniert und in einer Visualisierung dargestellt werden. Dafür werden die gesäuberten und mit der Spalte “name” versehenen Daten aus Schritt 4 verwendet und mit den Edge-Daten kombiniert. Die Spalten “modul” und “role” werden aus dem Algorithmus-Output erstellt und mit Werten befüllt. Da sich die Werte von “role” im “ordered factor”-Format vorliegen, müssen diese erst zu Zeichen konvertieren werden, um in der Visualisierung verwendbar zu sein.

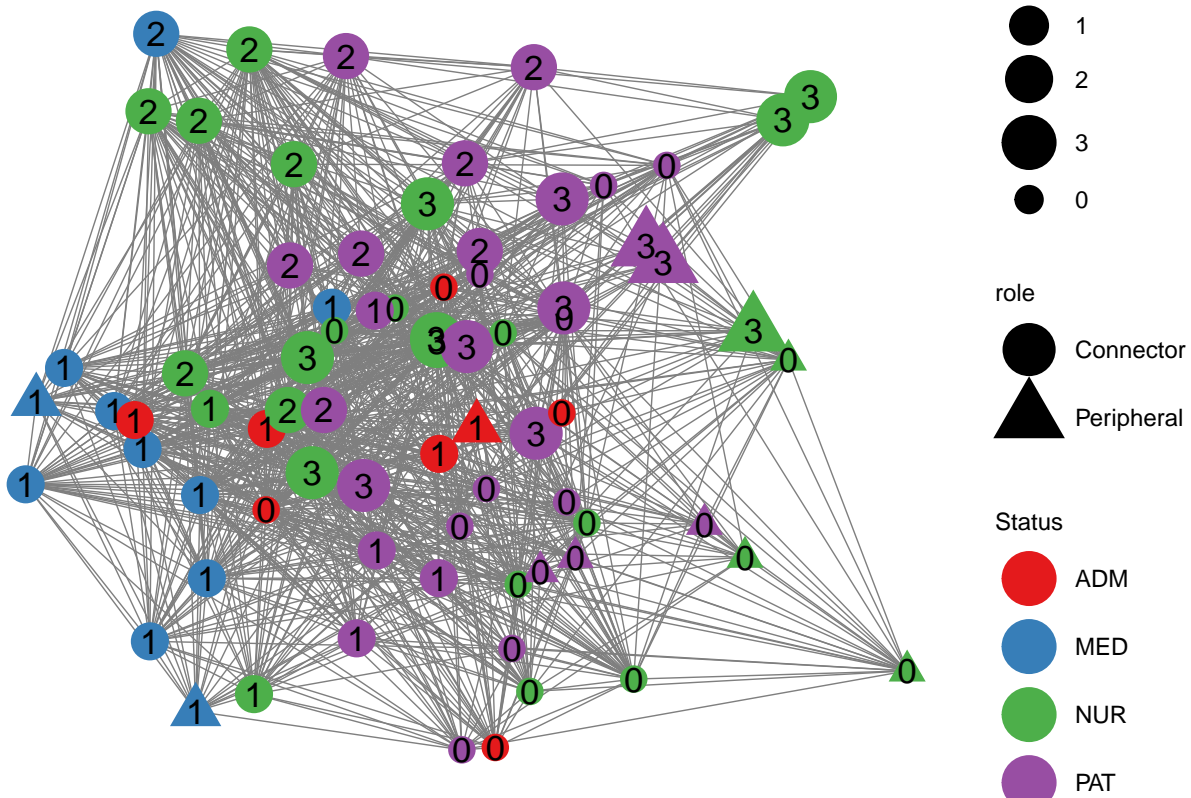
```
netPlot <- df2 %>% activate(nodes) %>% mutate(role <- merge(x = df2 , y = rnc[[1]] , c("name", "module"))
netPlot <- netPlot %>% as_tbl_graph() %>% activate(nodes) %>% mutate(role = unlist(lapply(role, as.char
```

7. Plot #2

Für den Plot des Netzwerks wird ggnet2 verwendet. Aus Gründen der Lesbarkeit und Vergleichbarkeit werden dieselben Visualisierungs-Eigenschaften verwendet wie in dem ersten Plot. Lediglich die Module wurden zusätzlich mit Labels visualisiert, da diese im Fokus stehen. Die Verteilung der Edges wird über den Modus “adj” realisiert, welcher auf der Adjazenzmatrix des Datensatzes basiert.

```
ggnet2(netPlot, mode = "adj", label = "module", node.size = "module", palette = "Set1", node.color = "S"
  ggtitle("Community-Struktur mit Modulen und Rollen")
```

Community–Struktur mit Modulen und Rollen



Interpretation der Visualisierungen

Zu Beginn wird die erste Visualisierung betrachtet. Hier ist klar zu sehen, dass es deutlich mehr „Connectors“ als „Peripherals“ gibt. Gerade Krankenschwestern weisen vergleichsweise sehr viele Begegnungen sowohl innerhalb als auch außerhalb ihres Netzwerks auf, was auch auf Ärzte zutrifft. Diese Berufsgruppen haben allgemein viele Kontakte und sind bis auf zwei als „Peripherals“ spezifiziert. Erstaunlicherweise haben auch einige Verwaltungsangestellte viele verschiedene Kontakte. Der „Within-community Degree“ ist bei Patienten allgemein nicht hoch und kommt nur sehr selten über 0. Dies lässt auf eher kleine Patientenzimmer schließen, welche nicht viele Betten haben. Die meisten Patienten werden jedoch mobil, und nicht an das Bett gebunden sein. Es gibt auch Ausreißer in den Daten. Eine Krankenschwester scheint fast ausschließlich Kontakt innerhalb ihrer Community gehabt zu haben. Aus dieser Darstellung lässt sich schließen, dass das größte Risiko, ein Superspreader zu sein, von den Krankenschwestern, den Ärzten, aber auch den Verwaltungsangestellten ausgeht. Diese Gruppen haben alle viele auch Community-übergreifende Kontakte. Diese These lässt sich auch durch die zweite Visualisierung bestätigen. In dem Netzwerk kann man klar erkennen, dass es keine zentralen Knotenpunkte gibt, sondern eine allgemein hohe Anzahl an Verknüpfungen auch zwischen den einzelnen Modulen besteht. Hier wird jedoch ersichtlich, dass die Ärzte fast exklusiv ein eigenes Modul bilden. Daraus lässt sich schließen, dass diese auch untereinander viel Kontakt, wie zum Beispiel Besprechungen, haben. Die Patient*innen und Krankenschwestern sowie -pfleger verteilen sich größtenteils über drei Module. Im Gegensatz dazu stehen die Verwaltungsangestellten, welche nur in zwei Modulen auftreten.

Conclusio

Wiederholung der Fragestellung

Zu Beginn wurde die Frage gestellt, ob gewisse Rollen in Krankenhäusern ein höheres Risiko haben, als ein Superspreader zu fungieren und ob die durch organisatorische und räumlich gegebene Bildung von Gemeinschaften einen Einfluss auf eine mögliche Verbreitung hat. Für die Beantwortung wurden zwei Visualisierungen erstellt: Das erste Diagramm ordnet die einzelnen Personen in Bezug auf ihre Kontakte innerhalb und außerhalb ihrer Community ein. Der zweite Plot zeigt das Netzwerk zusammen mit den errechneten Werten. Wertet man beide in Kombination aus können die folgenden Schlüsse gezogen werden.

Zusammenfassung der zentralen Ergebnisse

Krankenschwestern und Krankenpfleger haben das größte Risiko bei einer nicht entdeckten Infektion als Superspreader zu fungieren. Sie sind besonders gut Community-übergreifend vernetzt und können deshalb zu einer schnellen Verbreitung einer Krankheit beitragen. Dieses Ergebnis darf jedoch nicht pauschalisiert werden, da es auch Pflegepersonal gibt, deren Kontakte auf eine Community fokussiert sind. Das Gleiche gilt auch für Ärzt/innen. Durch ihre Tätigkeit haben Sie automatisch viel Kontakt mit anderen Personen. Hervorzuheben sind hier jedoch auch die Kontakte, welche innerhalb dieser Berufsgruppe stattfinden. Diese müssten im Extremfall unter besonderen Schutz abgehalten werden. Ein eher unerwartetes Ergebnis stellen die Verwaltungsangestellten dar. Auch sie haben teilweise ein sehr hohes epidemiologisches Potential. Ein wichtiges Takeaway wäre daher, diese bei Schutzkonzepten nicht zu übergehen, sondern sie aktiv mit in die Planung und Risikoanalyse einzubeziehen. Die Patienten sind größtenteils außerhalb von ihren Gruppen vernetzt. Dadurch haben sie zwar auch das Potential, einen Erreger von einem Modul in ein anderes zu übertragen, jedoch ein eher geringes Risiko der Übertragung innerhalb ihrer Community. Daraus lässt sich schließen, dass Personen wie Ärzte und Pfleger*innen die Hauptrollen in dem Communities spielen. Wenn diese Personen, zusammen mit den Angestellten der Verwaltung, sich und die Patienten schützen, kann eine Übertragung verhindert und die Patienten könnten geschützt werden. Eine weitere Möglichkeit wäre eine genaue Einteilung des Pflegepersonals auf bestimmte Zimmer. Dadurch würde sich der “Within-community Degree” deutlich erhöhen und der “Participatin Coefficient” gesenkt werden. Durch diese strukturellen Veränderungen würde eine Infektion nur eine Community affektieren und auch das Risiko der Ausbreitung zwischen den Modulen einschränken. Die Anwendung der genannten Maßnahmen kann das Risiko einer Verbreitung von gefährlichen Erregern in einem Krankenhaus deutlich verringern. Die in der Forschungsfrage beschriebene räumliche Trennung hat in dem Datensatz zu keiner deutlich sichtbaren Spaltung der Kontaktgruppen beigetragen. Es kann jedoch die Fragestellung hinsichtlich der Rollen definitiv beantwortet werden. Der nächste Schritt wäre ein Experiment mit den beschriebenen Maßnahmen und ein darauffolgender Vergleich der Datensätze. Zusammenfassend kann die Analyse erfolgreich abgeschlossen werden.

Limitationen, weiterführende Kommentare

Obwohl die Daten unter realen Bedingungen gesammelt wurden, lässt sich der Datensatz nur bedingt auf eine Situation wie die aktuelle SARS-CoV-2-Pandemie übertragen, da durch die gegebenen Regelungen eine wesentlich striktere Trennung der Patienten und der Angestellten vorgenommen wird, sowie ein höherer Schutz durch das Tragen von Masken oder Schutzkleidung gegeben ist. Durch die Ergebnisse kann aber gerade zu Beginn einer Infektionswelle die Ausbreitungsgeschwindigkeit durch strukturelle Änderungen gebremst werden.