

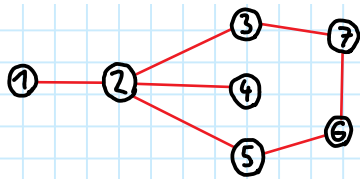
# Hausaufgabe 10

## Aufgabe 10.1 Netzwerkanalyse (40%)

Zeichnen Sie einen Graphen mit der folgenden Edgelist: (1,2),(2,4),(2,3),(2,5),(5,6),(3,7),(6,7)

Berechnen sie zusätzlich den Degree, die Betweenness, die Closeness, und den Clusteringkoeffizienten aller Knoten.

Wieviele Komponenten (connected components) hat der Graph? Berechnen sie dessen Dichte (density).



	Degree $k_i$	Betweenness	Closeness	Clustering Coefficient
1	1	0	$\frac{6}{5}$	/
2	4	$\frac{31}{6}$	$\frac{3}{4}$	0
3	2	$\frac{3}{5}$	$\frac{6}{13}$	0
4	1	0	$\frac{3}{5}$	/
5	2	$\frac{3}{5}$	$\frac{1}{2}$	0
6	2	$\frac{1}{2}$	$\frac{1}{2}$	0
7	2	$\frac{1}{2}$	$\frac{1}{2}$	0

$$C_B(n_2) = 5 \cdot \frac{1}{2} + 4 \cdot \frac{2}{3} = \frac{15}{6} + \frac{16}{6} = \frac{31}{6}$$

$$C_B(n_3) = C_B(n_5) = \frac{1}{2} + 2 \cdot \frac{1}{3} = \frac{2}{3} \dots$$

$$CC(1) = 6 \cdot (1 + 3 \cdot 2 + 2 \cdot 3)^{-1} = \frac{6}{13}$$

$$CC(2) = 6 \cdot (4 \cdot 1 + 2 \cdot 2)^{-1} = \frac{6}{8} = \frac{3}{4}$$

$$CC(3) = CC(5) = 6 \cdot (2 \cdot 1 + 4 \cdot 2)^{-1} = \frac{6}{10} = \frac{3}{5}$$

$$CC(4) = 6 \cdot (1 + 3 \cdot 2 + 2 \cdot 3)^{-1} = \frac{6}{13}$$

$$CC(6) = CC(7) = 6 \cdot (2 \cdot 1 + 2 \cdot 2 + 2 \cdot 3)^{-1} = \frac{1}{2}$$

Komponenten: 7

$$\text{Dichte } \rho = \frac{2 \cdot 7}{7 \cdot 6} = \frac{1}{3}$$

## Aufgabe 10.2 Netzwerkanalyse in R (40%)

Der visualisierter Graph ist in der Datei graphPlot.pdf zu finden.

Direkt in dieser Visualisierung ist zu erkennen, dass das Netzwerk aus interagierenden Proteinen einige Punkte besonders hoher Dichte besitzt und einige wenige Proteine eine besonders zentrale Rolle spielen. An den Rändern des Graphen befinden sich viele Proteine, die nur über eine Kante mit einem anderen Protein verbunden sind. In der Mitte finden sich dabei Proteine mit deutlich mehr angrenzenden Kanten.

Am auffälligsten ist das Protein APP, zu welchem besonders viele Kanten führen. Aber auch einige weitere Proteine wie NOTCH1 scheinen Ansatzpunkt für viele Kanten zu sein.

Eine Bestimmung und anschließende Sortierung der Closeness und Betweenness aller Knoten erlaubte dabei eine genauere Bestimmung der wichtigsten Knoten des Netzwerks - zur besseren Vergleichbarkeit wurde hierbei noch die Summe der Positionen eines Knoten in der Closeness- und Betweennesssortierung betrachtet.

Zuerst muss erwähnt werden, dass in der Berechnung für die Closeness wohl einige Fehler auftraten, da viele Knoten einen Wert von 1 aufwiesen, wobei sie selber sehr peripher am Netzwerk gelegen waren (siehe z.B. TOMM22). Die Sortierung in allen Kategorien zeigte das, was auch visuell schon erkenntlich war: Das sehr zentral gelegene Protein APP ist das relevanteste Protein in diesem Graphen. Genauso finden sich in beiden Kategorien sehr ähnliche Proteine weit oben in der Sortierung: So sind außerdem NOTCH1 und LRP1 in beiden Kategorien in den Top 5 und somit ebenfalls von essentieller Bedeutung für das Netzwerk. Dabei ist noch zu erwähnen, dass in beiden Kategorien APP einen herausstechend großen Wert annahm und damit mit sehr großer Sicherheit als das zentrale Protein identifiziert werden konnte, wohingegen sich die Werte für die folgenden Proteine nicht mehr so stark unterscheiden und damit wohl von ähnlicher Wichtigkeit für das Netzwerk sind.

## Aufgabe 10.3 Freundschaftsparadox (20%)

Das in dem Artikel beschriebene Freundschaftsparadox besagt, dass in einem sozialen Netzwerk die eigenen Freunde meist besser vernetzt (mehr Freunde haben) als man selbst. Dieser der Intuition widersprechende Umstand folgt aus der Eigenschaft von Netzwerken: Ein sehr viel verknüpfter Knoten ist mit sehr vielen Knoten verbunden und geht damit in der Betrachtung aller anderen Knoten mehrfach ein - wird also viel stärker gewichtet und bewirkt so eine Verschiebung des Mittelwertes in Richtung einer besseren Vernetztheit.

In der Biologie kann dieser Umstand zur Entwicklung von potenten Medikamenten oder Identifizierung von aussagekräftigen Biomarkern genutzt werden, denn wenn es gelingt, einen Punkt in einem Netzwerk (zum Beispiel PIN) zu finden, der deutlich mehr Vernetzungen als andere Knoten aufweist, so ist hier eine Einflussnahme oder Aussage über das System deutlich holistischer möglich als bei einem zufällig ausgewählten Knoten.