Skript zur Creative Challenge
Stochastische Irrfahrten auf Graphen

Prof. Dr. Noemi Kurt, Goethe-Universität Frankfurt

Main Math Challenge, 7. Oktober 2023

In diesem Skript und den zugehörigen Videos wird ein Thema aus der Stochastik eingeführt, welches zur Bearbeitung der Creative Challenge verwendet werden kann und soll. Die Aufgabe zur Creative Challenge wird an der Main Math Challenge gestellt werden. Wir empfehlen, vor der Main Math Challenge die Videos zu schauen und dieses Skript durchzuarbeiten. Die Übungsaufgaben im Skript sind freiwillig und nicht Bestandteil der Aufgabe bei der Creative Challenge.

1 Graphen

Als erstes definieren wir formal, was ein *Graph* ist. Siehe dazu auch Abbildung 1 auf der nächsten Seite.

Definition 1 (Graph)**.** Ein Graph ist ein Objekt bestehend aus einer Menge von Knoten (manchmal auch Ecken genannt) und einer Menge von Kanten, wobei eine Kante jeweils zwei Knoten bzw. Ecken miteinander verbindet.

Wir schreiben V für die Menge der Knoten, und E für die Menge der Kanten. Hier nehmen wir stets an, dass wir N verschiedene Knoten haben, also $|V| = N$, wobei $|V|$ die Anzahl der Elemente der Menge V bezeichnet. Wir nummerieren diese Knoten von 1 bis N, bzw. wir wählen $V = \{1, ..., N\}$ als Knotenmenge. Es sei stets $N \geq 2$. Für zwei Knoten $i, j \in V$ schreiben wir $i \sim j$, falls es zwischen i und j eine Kante gibt, falls also i und j *Nachbarknoten* sind. In Abbildung 1 sind z.B. 14 und 12 Nachbarknoten, also $12 \sim 14$. Jedoch sind 10 und 14 keine Nachbarknoten.

Zusätzlich machen wir einige Annahmen an den Graphen:

(A1) Jede Kante verbindet genau zwei verschiedene Knoten miteinander (also keine "Schlaufen" von einem Knoten zu sich selbst)

(A2) Zwischen jedem Paar von Knoten gibt es höchstens eine Kante (also keine Mehrfachkanten)

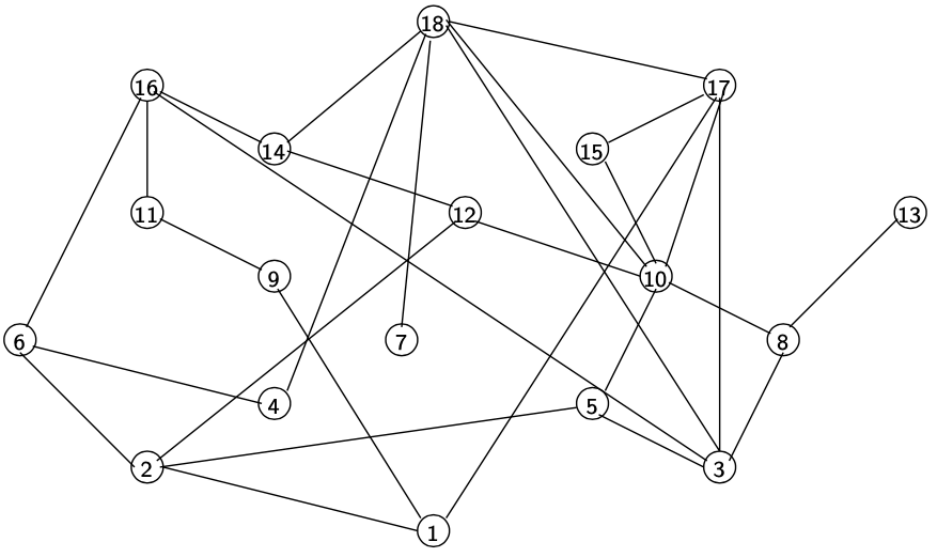

Abbildung 1: Ein Beispiel für einen Graphen mit 18 Knoten. Die Knoten sind hier als nummerierte Kreise dargestellt, die Kanten als Linien zwischen zwei Knoten.

(A3) Der Graph ist *zusammenhängend*, es gibt also zwischen jedem Paar von Knoten $i, j \in V$ eine Abfolge von Kanten und Knoten, welche einen "Pfad" von i nach j bildet (wenn man die Kanten sukzessive durchläuft).

Der Graph in Abbildung 1 erfüllt alle diese Voraussetzungen. So gibt es z.B. einen Pfad von aufeinanderfolgenden Kanten von 1 nach 16, welche keine Nachbarknoten sind, indem man die Zwischenknoten 9 und 11 verwendet. Analog findet man Pfade für jedes beliebige Paar von Knoten.

Definition 2 (Grad eines Knoten)**.** Jeder Knoten $i \in V$ eines Graphen hat einen *Grad* $d_i \in \mathbb{N}$, das ist die Anzahl Kanten, die mit diesem Knoten verbunden sind. In anderen Worten, d_i ist die Anzahl Nachbarknoten von i.

In Abbildung 1 hat beispielsweise Knoten 13 nur einen Nachbarn, also $d_{13} = 1$, hingegen hat Knoten 17 insgesamt 5 Nachbarn bzw. Kanten, also $d_{17} = 5$.

2 Irrfahrt auf einem Graphen

In der Stochastik interessiert man sich oft für *Irrfahrten auf Graphen*. Das bedeutet folgendes: Man stellt sich vor, dass man über den Graphen hüpfen kann, indem man von Knoten zu Knoten springt, allerdings nur dann, wenn zwischen den Knoten eine Kante ist. Dabei notiert man sich nach jedem Sprung die Nummer des Knoten, an dem man sich gerade befindet. Von Knoten i ausgehend sucht man sich das Ziel des nächsten Sprunges *zufällig* aus, und zwar so, dass man mit gleicher Wahrscheinlichkeit zu einem der Knoten springt, die mit i verbunden sind. Davon gibt es nach Definition d_i Stück, zu jedem davon springt man also mit Wahrscheinlichkeit $1/d_i$. Etwas formaler: Wir schreiben

$$p_{i,j} = \frac{1}{d_i} \quad \text{falls } i \sim j$$

also $p_{i,j}$ ist die Wahrscheinlichkeit, von Knoten i zu Knoten j zu springen, falls diese durch eine Kante miteinander verbunden sind. Falls j keine gemeinsame Kante mit i hat, schreiben wir $p_{i,j} = 0$.

Übungsaufgabe 1. *Man mache sich bewusst, dass $p_{i,j}$ nicht gleich $p_{j,i}$ sein muss. Jedoch gilt $p_{i,j} > 0 \Leftrightarrow p_{j,i} > 0$.*

Definition 3 (Irrfahrt auf einem Graphen)**.** Eine einfache symmetrische Irrfahrt auf einem Graphen mit Knotenmenge $V = \{1, ..., N\}$ und Kantenmenge E ist eine zufällige Abfolge von Knoten des Graphen. Dabei startet man in einem vorgegebenen Knoten. In jedem Schritt wird der nächste Knoten folgendermaßen bestimmt: Falls man sich aktuell im Knoten mit der Nummer i befindet, springt man mit Wahrscheinlichkeit $p_{i,j}$ im nächsten Schritt nach j.

Startet man beispielsweise im Graphen aus Abbildung 1 in Knoten Nr. 6, so hat man $d_6 = 3$ Möglichkeiten für den nächsten Sprung, mit Wahrscheinlichkeit 1/3 springt man also nach 2, mit Wahrscheinlichkeit 1/3 nach 4 und mit Wahrscheinlichkeit 1/3 nach 16. Springt man nach 4, hat man danach $d_4 = 2$ Möglichkeiten, man springt also mit Wahrscheinlichkeit 1/2 nach 18 und mit Wahrscheinlichkeit 1/2 zurück nach 6. Von Knoten 13 aus hat man nur eine Möglichkeit, landet man irgendwann dort, springt man also im nächsten Schritt direkt zurück nach 8.

Eine solche Irrfahrt ist ein Beispiel für einen *stochastischen Prozess*, also eine Folge von zufälligen Werten. Dabei notiert man sich die zufällig entstandene Folge der Nummern der Knoten, die in diesem Verfahren besucht werden.

3 Gewicht eines Knoten und Rückkehrzeiten der Irrfahrt

Wir definieren nun eine letzte wichtige Größe für unsere Aufgabe.

Definition 4 (Gewicht eines Knoten)**.** Für jeden Knoten i unseres Graphen sei

$$\pi_i := \frac{d_i}{d_1 + ... + d_N}$$

das *Gewicht* des Knoten i.

Übungsaufgabe 2. *Zeigen Sie, dass unter unseren Annahmen gilt $0 < \pi_i < 1$ für alle Knoten i, und $\pi_1 + ... + \pi_N = 1$.*

Man kann wegen Übungsaufgabe 2 also das Gewicht π_i auch als Wahrscheinlichkeit interpretieren. Nach Definition ist π_i proportional zu d_i, die Proportionalitätskonstante ist $1/(d_1 + ... + d_N)$. Das Gewicht π_i ist also ein Maß dafür, ob ein Knoten $i \in V$ im Vergleich zu den anderen Knoten viele Kanten oder wenige hat: Je mehr Kanten (bzw. Nachbarknoten) i hat, desto höher ist sein Gewicht. Intuitiv sind Knoten mit einem höheren Gewicht "wichtiger" als solche mit einem niedrigen. Etwas genauer: Je höher das Gewicht ist, desto mehr Kanten hat der Knoten, desto bedeutender sein Einfluss im Graphen. Und: Desto häufiger wird der Knoten von einer Irrfahrt auf dem Graphen auf lange Sicht besucht - dazu unten noch etwas mehr.

Satz 1. *Für die oben definierte Irrfahrt auf einem Graphen gelten unter unseren Voraussetzungen*

(a) $\pi_i p_{i,j} = \pi_j p_{j,i}$ *für alle* $i, j \in V$

(b) $\pi_1 p_{1,j} + \pi_2 p_{2,j} + ... + \pi_N p_{N,j} = \pi_j$ *für alle* $j \in V$

Beweis. Aussage (a) folgt direkt aus den Definitionen mittels Kürzen bzw. Erweitern: Falls $i \sim j$,

$$\pi_i p_{i,j} = \frac{1}{d_i} \frac{d_i}{d_1 + ... + d_N} = \frac{1}{d_1 + ... + d_N} = \frac{1}{d_j} \frac{d_j}{d_1 + ... + d_N} = \pi_j p_{j,i}.$$

Andernfalls ist $p_{i,j} = p_{j,i} = 0$, und die Aussage stimmt ebenfalls. Aussage (b) erhält man, indem man in Aussage (a) auf beiden Seiten die Summe über i bildet:

$$\pi_1 p_{1,j} + ... + \pi_N p_{N,j} = \pi_j p_{j,1} + ... + \pi_j p_{j,N} = \pi_j (p_{j,1} + ... + p_{j,N}).$$

Nun müssen wir uns nur daran erinnern, dass $p_{j,i}$ die Wahrscheinlichkeit ist, von j nach i zu springen. Wenn wir nun alle i zusammenzählen, ist also $p_{j,1} + ... + p_{j,N}$ einfach die Wahrscheinlichkeit, von j aus in irgendeinen beliebigen Knoten zu springen. Und da man auf jeden Fall irgendwohin springt, ist diese Wahrscheinlichkeit gleich 1, also

$$\pi_1 p_{1,j} + ... + \pi_N p_{N,j} = \pi_j (p_{j,1} + ... + p_{j,N}) = \pi_j \cdot 1 = \pi_j.$$

□

Dieser Satz besagt, dass man die Gewichte π_i auch als eine Art *Gleichgewicht* der Irrfahrt interpretieren kann. Im Video wird das etwas ausgeführt.

Die Bedeutung von π liegt darin, dass die Gewichte das Langzeitverhalten der Irrfahrt beschreiben. Wir schreiben $\mathbb{E}_i[T_i]$ für die durchschnittliche Anzahl Sprünge, nach der die Irrfahrt bei Start in i zum ersten Mal wieder in Knoten i landet, also die erwartete erste Rückkehrzeit zum Startknoten i. Dann kann man folgenden Satz beweisen:

Satz 2. *Für die Irrfahrt auf einem Graphen mit N Knoten gilt unter unseren Annahmen*

$$\mathbb{E}_i[T_i] = \frac{1}{\pi_i} = \frac{d_1 + ... + d_N}{d_i} \qquad \textit{für alle } i \in V.$$

Dieser Satz sagt also aus, dass die erwartete erste Rückkehrzeit zum Startpunkt gleich dem Kehrwert des Gewichts π_i dieses Knotens ist. Knoten mit großem Gewicht werden also relativ bald wieder besucht, Knoten mit kleinem Gewicht erst nach einer längeren Zeit. Wir verzichten auf einen formalen Beweis dieser Aussagen, liefern aber im Video einen Ansatz einer anschaulichen Begründung. Der formale Beweis ist nicht sehr schwer, aber etwas lang für diese Einführung.

Zum Schluss möchten wir noch einen kleinen Ausblick geben, der jedoch für die Bearbeitung der Creative Challenge nicht benötigt wird. Wir schreiben $\mathbb{P}_j(X_n = i)$ für die Wahrscheinlichkeit, dass die Irrfahrt bei Start in j nach n Sprüngen in Knoten i gelandet ist. Im Mathematikstudium lernt man normalerweise, dass mit Hilfe eines mathematisch interessanten *Kopplungsarguments* bewiesen werden kann, dass in sehr vielen Situationen (beispielsweise für die Irrfahrt auf dem Graphen aus Abbildung 1) gilt:

$$\lim_{n \to \infty} \mathbb{P}(X_n = i) = \pi_i.$$

Das bedeutet, dass die Wahrscheinlichkeit, sich nach n Schritten in Knoten i zu befinden, wenn man dabei n gegen unendlich gehen lässt, genau dem Gewicht des Knoten i entspricht, und zwar egal, wo man gestartet ist. Auf lange Sicht geben also die Gewichte exakt die Wahrscheinlichkeitsverteilung des Aufenthaltsortes wieder. Knoten mit großem Gewicht werden also häufiger besucht als Knoten mit kleinem Gewicht. Diese Aussage ist für viele Anwendungen von großer Bedeutung. Beispielsweise basiert auch der Google-Algorithmus PageRank darauf. Er basiert darauf, dass man Websites als Knoten in einem riesigen Graphen interpretiert, und Links zwischen Websites als Kanten des Graphen. Durch Klicken auf Links springt man von einem Knoten zum nächsten. Wenn man die Gewichte (ungefähr) kennt, geben sie ein Maß für die Wichtigkeit der entsprechenden Websites an, so dass Google die Suchergebnisse nach Wichtigkeit ordnen kann.