

# Markov-Ketten und Google's Page-Rank

1. Wir geben einen kurzen Überblick über die Arbeitsweise von **Suchmaschinen** für das Internet.
  - ▶ Eine Suchmaschine erwartet als Eingabe ein Stichwort oder eine Liste von Stichworten
  - ▶ und gibt als Ausgabe eine Liste von Links auf möglichst informative Webseiten zu diesen Stichworten. Die Liste soll so sortiert sein, dass  
**die informativsten Links am „weitesten oben“ stehen.**
2. Gleichzeitig geben wir eine Einführung in **Markov-Ketten**,  
einem wichtigen Werkzeug in der Modellierung einfacher stochastischer Prozesse.

Die von Suchmaschinen zu bewältigenden Datenmengen sind immens!

In 2018 gab es

- ca. 1,8 Milliarden Websites (von denen ca. 200 Millionen regelmäßig aktualisiert werden) mit ca. 1 Milliarde neuer Websites seit 2016,
- ca. 4 Milliarden Internetnutzer weltweit
- und ca. 2 Billionen, also  $10^{12}$  Suchanfragen pro Jahr auf Google allein.

Und dann müssen Suchanfragen auch noch in „Echtzeit“ beantwortet werden.

Die zentrale Aufgabe:

**Bewerte den Informationsgehalt der Webseiten**

in Bezug auf die jeweilige Kombination der Suchbegriffe!

# Die Architektur von Suchmaschinen

Anfragen für einen sich rasant ändernden Suchraum gigantischer Größe sind ohne merkliche Reaktionszeit zu beantworten.

- (1) **Web-Crawler** durchforsten das Internet, um neue oder veränderte Webseiten zu identifizieren.
- (2) Die von den Crawlern gefundenen Informationen werden in einer komplexen **Datenstruktur** gespeichert, um bei Eingabe von Suchbegriffen in „Echtzeit“ alle relevanten Webseiten ermitteln zu können.
- (3) **Bewerte die Webseiten**  
*hinsichtlich ihrer Relevanz für mögliche Suchbegriffe wie auch hinsichtlich ihrer **generellen** Bedeutung im Internet.*

# Datenstrukturen für Suchmaschinen

# Die Datenstruktur: Index und invertierter Index

1. Im **Index** werden alle vom Crawler gefundenen Webseiten  $w$  gespeichert:
  - ▶ URL (d.h. die Adresse) und Inhalt von  $w$ .
  - ▶ Der Inhalt von  $w$  wird analysiert: Alle vorkommenden Worte werden in Häufigkeit und Sichtbarkeit (Vorkommen in Überschriften, Schriftgröße etc.) erfasst.
  - ▶ Die auf  $w$  zeigenden Hyperlinks werden ebenfalls analysiert:
    - ★ Welche Begriffe tauchen in der Beschriftung des Links auf?
    - ★ Wie prominent ist der Link platziert?
2. Aus dem Index wird der **invertierte Index** erzeugt, der zu jedem möglichen Suchbegriff eine Liste aller Webseiten enthält, die den Suchbegriff enthalten.
  - ▶ Für jede in der Liste auftauchende Webseite  $w$  wird die Sichtbarkeit des Begriffs innerhalb von  $w$  und innerhalb der auf  $w$  zeigenden Seiten aufgeführt.
  - ▶ Mit diesen Zusatzinformationen und mit Hilfe ihrer

## „grundsätzlichen“ Bedeutung

wird die Seite  $w$  in die Liste eingereiht.

- ★ Wie die Einreihung erfolgt, ist Betriebsgeheimnis der Suchmaschinenbetreiber.

Wie bestimmt man **die grundsätzliche Bedeutung einer Webseite**?

# Der Page-Rank einer Webseite



# Peer Review: Die grundsätzliche Bedeutung einer Webseite

Im Ansatz des „**Peer Review**“ wird die folgende Annahme gemacht:

Wenn eine Webseite  $i$  einen Link auf eine Webseite  $j$  enthält, dann

1. gibt es eine inhaltliche Beziehung zwischen beiden Webseiten, und
2. der Autor der Webseite  $i$  hält die Informationen auf Webseite  $j$  für wertvoll.

Der **Web-Graph**,

also die Link-Struktur des Internets

spielt im Peer-Review eine besondere Rolle. Zur Erinnerung:

- Die Webseiten sind Knoten und
- die Hyperlinks sind die gerichteten Kanten des Webgraphen.

# Page-Rank: Notation

Um die „grundlegende Bedeutung“ einer Webseite zu messen, berücksichtigt der Page-Rank nur die Link-Struktur des Internets, nicht aber den Inhalt der Seite.

$W_{\text{EB}} = (V, E)$  bezeichne im Folgenden den Web-Graphen.

- Der Einfachheit halber nehmen wir an, dass die Webseiten mit den Zahlen  $1, \dots, n$  durchnummeriert sind, und dass  $V = \{1, 2, \dots, n\}$  gilt.
- Für jeden Knoten  $i \in V$  ist

$$a_i := \text{Aus-Grad}_{W_{\text{EB}}}(i)$$

der Ausgangsgrad von  $i$  in  $W_{\text{EB}}$ , also die Anzahl der Hyperlinks, die von der Webseite  $i$  auf andere Webseiten verweisen.

- Für eine Webseite  $j \in V$  schreiben wir  $\text{Vor}_{W_{\text{EB}}}(j)$ , um die Menge aller Webseiten zu bezeichnen, die einen Link auf  $j$  enthalten, d.h.

$$\text{Vor}_{W_{\text{EB}}}(j) = \{i \in V : (i, j) \in E\}.$$

Die Elemente in  $\text{Vor}_{W_{\text{EB}}}(j)$  heißen (direkte) Vorgänger von  $j$ .

# Page-Rank mittels Peer Review

Wir messen die „grundlegende Bedeutung“ einer Webseite  $j$  durch die Zahl

$$PR_j,$$

den Page-Rank von  $j$ .

*Der Wert  $PR_j$  soll das „Renommee“ der Webseite  $j$  widerspiegeln:  
 $PR_j$  soll umso größer sein, je höher das Renommee der Webseite  $j$  ist.*

Wann sollte  $PR_j$  groß sein?

*Wenn genügend viele Webseiten mit großem Page-Rank auf  $j$  zeigen!*

Wir fordern, dass eine Webseite  $i$  ihren Page-Rank an alle Webseiten  $j$ , auf die  $i$  zeigt, zu gleichen Teilen „vererbt“.

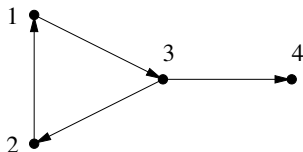
$$PR_j = \sum_{i \in \text{Vor}_{\text{WEB}}(j)} \frac{PR_i}{a_i}.$$

Wir erhalten ein **Gleichungssystem**: Eine Gleichung für jeden Knoten  $j$ .

# Schauen wir mal, was passiert

**Senken**, also Knoten vom Ausgangsgrad 0, vererben ihren Page-Rank nicht.  
Ist das problematisch?

Betrachte den folgenden „Webgraphen“  $W_{EB} = (V, E)$ :



Die einzigen Page-Rank Werte, die die Gleichungen

$$PR_j = \sum_{i \in \text{Vor}_{W_{EB}}(j)} \frac{PR_i}{a_i}.$$

erfüllen, sind  $PR_1 = PR_2 = PR_3 = PR_4 = 0$  und diese Werte sollen die

„grundlegende Bedeutung“

der 4 Seiten widerspiegeln?

Um Senken loszuwerden:

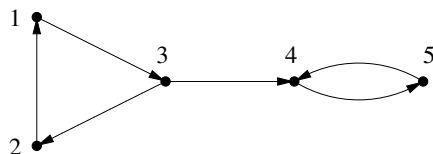
1. Füge von jeder Senke Kanten zu *allen* Knoten hinzu oder
2. lösche rekursiv(!) alle Senken, bis ein Graph ohne Senken übrig bleibt.

Welche Transformation auch immer durchgeführt wird: Der Web-Graph wird durch einen gerichteten Graphen  $WEB = (V, E)$  **ohne Senken** repräsentiert.

# Haben wir das Problem gelöst?

Was passiert, wenn bestimmte Knoten nur unter sich verbunden sind, aber keine Kante zu einem anderen Knoten des Graphen  $W_{EB}$  besitzen?

Betrachte den folgenden Graphen  $W_{EB} = (V, E)$ :



Man kann sich leicht davon überzeugen, dass Page-Rank Werte die Gleichung

$$PR_j = \sum_{i \in \text{Vor}_{W_{EB}}(j)} \frac{PR_i}{a_i}.$$

genau dann erfüllen, wenn  $PR_1 = PR_2 = PR_3 = 0$  und  $PR_4 = PR_5$  gilt.

Ist das jetzt gut oder schlecht?

Gehört unser Ansatz „in die Tonne“  
oder stimmt die Grundidee?

# Woran liegt's?

## (a) Der Ausweg?

- ▶ Füge Kanten von einer Webseite  $i$  zu **allen** anderen Webseiten ein, „**dämpfe**“ aber den Beitrag der neuen Seiten mit dem Faktor  $1 - d$  für  $0 \leq d \leq 1$ .
- ▶ Die alten Kanten, also die Hyperlinks, werden mit dem Faktor  $d$  gedämpft.

## (b) Man stelle sich vor, dass Renommee im Gesamtumfang von 1 auf die Webseiten verteilt wird:

- ▶ Ein **Grundeinkommen** im Gesamtumfang von  $1 - d$  wird gleichmäßig auf die einzelnen Seiten verteilt,
- ▶ der **Verdienst** im Gesamtumfang von  $d$  wird mit Peer Review erworben.

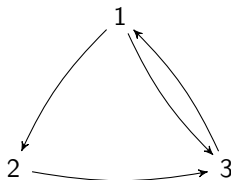
## (c) Ein Tupel $PR = (PR_1, \dots, PR_n) \in \mathbb{R}^n$ hat die **Page-Rank-Eigenschaft** bezüglich $d$ , wenn für alle $j \in V$ gilt:

$$PR_j = \underbrace{(1 - d) \cdot \frac{1}{n}}_{\text{anteiliges Grundeinkommen}} + \underbrace{d \cdot \sum_{i \in \text{Vor}_{\text{WEB}}(j)} \frac{PR_i}{a_i}}_{\text{durch Peer Review erworbener Verdienst}} .$$



# Der neue Page-Rank: Ein Beispiel für $d = \frac{1}{2}$

Betrachte den folgenden Graphen  $WEB = (V, E)$ :



Wir suchen ein Tupel  $PR = (PR_1, PR_2, PR_3)$  von reellen Zahlen mit der Page-Rank-Eigenschaft bezüglich  $d = 1/2$ :

- (1)  $PR_1 = \frac{1}{2 \cdot 3} + \frac{1}{2} \cdot \frac{PR_3}{1}$ ,
- (2)  $PR_2 = \frac{1}{2 \cdot 3} + \frac{1}{2} \cdot \frac{PR_1}{2}$
- (3)  $PR_3 = \frac{1}{2 \cdot 3} + \frac{1}{2} \cdot \left( \frac{PR_1}{2} + \frac{PR_2}{1} \right)$ .

Wir lösen das lineare Gleichungssystem und erhalten

$$PR_1 = \frac{14}{39}, \quad PR_2 = \frac{10}{39}, \quad PR_3 = \frac{15}{39}.$$

# Page-Rank: Die wichtigen Fragen

Wir müssen ein lineares Gleichungssystem lösen.

- (a) Ist das Gleichungssystem überhaupt lösbar und wenn ja, ist die Lösung eindeutig?
- (b) Und wie soll man, bitte schön, ein Gleichungssystem mit mehreren Milliarden Zeilen und Spalten lösen?
  - ▶ Unsere Rechner sind mittlerweile so mächtig: kein Problem mit Gaußscher Eliminierung!
  - ▶ **Denkste!** Ein Gleichungssystem dieser Dimension können wir auch mit allen Rechnern dieser Welt nicht knacken, wenn ...  
...wir eine Gaußsche Eliminierung ausführen müssen.
- (c) Und selbst wenn es genau eine Lösung PR gibt und wir diese Lösung irgendwie bestimmen können:

Gibt  $PR_i$  das Renommee der Webseite  $i$  wieder?

Wir betrachten gleich einen rivalisieren Ansatz mit dem Page-Rank aus

**Perspektive des Zufallssurfers.**

# Der Zufallssurfer

Ein Tupel  $\pi \in \mathbb{R}^n$  heißt eine **Verteilung** (auf  $\{1, \dots, n\}$ ), falls  $\pi$  die beiden folgenden Eigenschaften hat:

- $\pi_i \geq 0$  für alle  $i \in \{1, \dots, n\}$  und
- $\sum_{i=1}^n \pi_i = 1$  gilt.

Wir benutzen Verteilungen, um **Irrfahrten** (engl: Random Walks) im Webgraphen  $\text{WEB} = (V, E)$  zu beschreiben:

Dazu legen wir für jeden Knoten  $i \in V$  eine Verteilung

$$(p_{i,1}, \dots, p_{i,n})$$

fest, so dass – zu jedem Zeitpunkt der Irrfahrt –  $p_{i,j}$  die Wahrscheinlichkeit ist, in einem Schritt von Knoten  $i$  zum Knoten  $j$  zu springen.

# Der Zufallssurfer

Mit welcher Wahrscheinlichkeit  $p_{i,j}$  soll unser Surfer  
vom Knoten  $i$  auf den Knoten  $j$   
springen?

Wenn der Zufallssurfer auf einer Webseite  $i$  ist, so wählt er

- die **Option „wildes Hüpfen“** mit Wahrscheinlichkeit  $(1 - d)$ .
  - ▶ Daraufgehend wähle eine der  $n$  Webseiten mit Wahrscheinlichkeit  $\frac{1}{n}$  aus.
- die **Option „Webgraph“** mit Wahrscheinlichkeit  $d$ .
  - ▶ Daraufgehend wähle einen der  $a_i = \text{Aus-Grad}_{\text{WEB}}(i)$  ausgehenden Links mit Wahrscheinlichkeit  $\frac{1}{a_i}$  aus.

Setze

$$p_{i,j} := \begin{cases} \frac{1-d}{n} + d \cdot \frac{1}{a_i} & \text{falls } (i,j) \text{ eine Kante von WEB ist,} \\ \frac{1-d}{n} & \text{sonst.} \end{cases}$$

( $d$  ist der Dämpfungsfaktor und  $a_i = \text{Aus-Grad}_{\text{WEB}}(i)$ .)

# Die Übergangsmatrix

# Die Übergangsmatrix $P_d(\text{WEB})$

wird definiert durch

$$P_d(\text{WEB}) := \begin{pmatrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & & \vdots \\ p_{n,1} & \cdots & p_{n,n} \end{pmatrix}$$

wobei

$$p_{i,j} = \begin{cases} \frac{1-d}{n} + d \cdot \frac{1}{a_i} & \text{falls } (i,j) \text{ eine Kante von WEB ist,} \\ \frac{1-d}{n} & \text{sonst.} \end{cases}$$

Für jedes  $i \in \{1, \dots, n\}$  ist die  $i$ te Zeile

$$(p_{i,1}, \dots, p_{i,n})$$

tatsächlich eine Verteilung, denn

$$\sum_{j=1}^n p_{i,j} = \sum_{j=1}^n \frac{1-d}{n} + \sum_{j:(i,j) \in E} \frac{d}{a_i} = n \cdot \frac{1-d}{n} + a_i \cdot \frac{d}{a_i} = 1 - d + d = 1.$$

Eine  $n \times n$  Matrix

$$P = \begin{pmatrix} P_{1,1} & \cdots & P_{1,n} \\ \vdots & & \vdots \\ P_{n,1} & \cdots & P_{n,n} \end{pmatrix}$$

heißt **stochastisch**, wenn

- (1)  $P_{i,j} \geq 0$  für alle  $i, j \in \{1, \dots, n\}$  und
- (2) für jede Zeile  $i \in \{1, \dots, n\}$  gilt:  $\sum_{j=1}^n P_{i,j} = 1$ .

Die Matrix  $P$  ist also genau dann stochastisch, wenn jede Zeile eine Verteilung ist

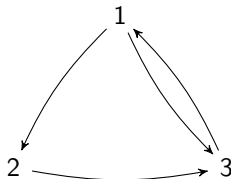


Die Matrix  $P_d(\text{WEB})$  der Übergangswahrscheinlichkeiten ist stochastisch.



# Die Übergangsmatrix

Für den Wert  $d = \frac{1}{2}$  und den Graphen WEB



ist beispielsweise  $P_{1/2}(\text{WEB})_{1,1} = \frac{1}{6}$ ,

$P_{1/2}(\text{WEB})_{1,2} = \frac{1}{6} + \frac{1}{4} = \frac{5}{12} = P_{1/2}(\text{WEB})_{1,3}$  und  $P_{1/2}(\text{WEB})_{2,3} = \frac{1}{6} + \frac{1}{2} = \frac{2}{3}$ .

Die vollständige Übergangsmatrix ist

$$P_{1/2}(\text{WEB}) = \begin{pmatrix} \frac{1}{6} & \frac{5}{12} & \frac{5}{12} \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

# „Peer Review“ versus „Zufallssurfer“

1. Der Peer-Review Ansatz: Ein Tupel  $PR = (PR_1, \dots, PR_n) \in \mathbb{R}^n$  hat die **Page-Rank-Eigenschaft** bezüglich  $d$ , wenn für alle  $j \in V$  gilt:

$$PR_j = \frac{1-d}{n} + d \cdot \sum_{i \in \text{Vor}_{\text{WEB}}(j)} \frac{PR_i}{a_i}.$$

2. Der **Zufallssurfer** führt eine zufällige **Irrfahrt** durch: Er springt mit Wahrscheinlichkeit  $p_{i,j}$  von der Webseite  $i$  zur Webseite  $j$ , wobei

$$p_{i,j} = \begin{cases} \frac{1-d}{n} + d \cdot \frac{1}{a_i} & \text{falls } (i,j) \text{ eine Kante von WEB ist,} \\ \frac{1-d}{n} & \text{sonst.} \end{cases}$$

Als Konkurrenz zum Page-Rank PR:

$PR_j^*$  = die relative Häufigkeit mit der Seite  $j$  in einer unendlich langen Irrfahrt besucht wird.

3.  $P_d(\text{WEB}) := \begin{pmatrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & & \vdots \\ p_{n,1} & \cdots & p_{n,n} \end{pmatrix}$  heißt Übergangsmatrix des Zufallssurfers.

# Markov-Ketten, das mathematische Modell für Irrfahrten

Der Zufallssurfer springt zufällig im Webgraphen herum: Er führt eine Irrfahrt durch. Was müssen wir über Irrfahrten wissen?

Den Graphen und die Übergangsmatrix.

$G = (V, E)$  sei ein gerichteter Graph mit Knotenmenge  $V = \{1, \dots, n\}$ . Die (homogene) **Markov-Kette**  $\mathcal{M}$  wird beschrieben durch das Paar

$$\mathcal{M} := (G, P)$$

mit dem Graphen  $G$  und der Übergangsmatrix  $P$ .

- (a)  $G$  hat keine Senke, d.h.  $\text{Aus-Grad}_G(v) > 0$  gilt für alle Knoten  $v$  von  $G$ .
- (b) Die Matrix  $P$  ist eine stochastische Matrix mit  $n$  Zeilen und  $n$  Spalten. Es ist  $P_{i,j} = 0$  genau dann, wenn  $(i, j)$  keine Kante in  $G$  ist.

Die Knoten von  $G$  nennt man häufig auch **Zustände**.

Wie stellt man eine Markov-Kette grafisch dar?

# Die von einer Markov-Kette erzeugte Irrfahrt

Sei  $\mathcal{M} = (G, P)$  eine Markov-Kette mit dem Graphen  $G = (V, E)$  und  $\pi$  eine auf  $V$  definierte Wahrscheinlichkeitsverteilung.

$\mathcal{M}$  **beginne** mit Wahrscheinlichkeit  $\pi_i$  im Zustand  $i \in V$ .

**Welche Irrfahrt wird von  $\mathcal{M}$  erzeugt?**

1.  $X_k$  bezeichne den von  $\mathcal{M}$  nach  $k$  Schritten erreichten Zustand.
  - ▶ Es gilt  $X_0 = i$  mit Wahrscheinlichkeit  $\pi_i$ .
  - ▶  $X_k$  ist eine Zufallsvariable.
2. Die Kette  $\mathcal{M} = (G, P)$  erzeugt die unendlich lange Irrfahrt in  $G$

$$X_0, X_1, X_2, \dots, X_k, \dots$$

1. **Wichtige Eigenschaft:** Der zum Zeitpunkt  $k > 0$  angenommene Zustand  $X_k$  hängt nur vom Vorgänger-Zustand  $X_{k-1}$  ab.
2. Mit welcher Wahrscheinlichkeit gilt  $X_1 = j$ ? D.h. mit welcher Wahrscheinlichkeit befindet sich die Irrfahrt nach **einem** Schritt im Zustand  $j$ ?

# Ein Schritt einer Markov-Kette: Das Vektor-Matrix Produkt

## (a) Um das **Vektor-Matrix Produkt**

$$y = x^T \cdot A$$

für eine  $n \times m$  Matrix  $A$  reeller Zahlen und ein Tupel  $x \in \mathbb{R}^n$  zu berechnen:

- ▶ interpretiere  $x$  als Zeilenvektor, den man dann nacheinander mit allen Spalten von  $A$  multiplizieren muss, also

$$y_i = \sum_{\ell=1}^n x_{\ell} \cdot a_{\ell,i}.$$

## (b) Um das **Matrizenprodukt**

$$C = A \cdot B$$

für eine  $n \times m$  Matrix  $A$  und eine  $m \times r$  Matrix  $B$  zu berechnen:

- ▶ Multipliziere die  $i$ te Zeile von  $A$  mit der  $j$ ten Spalte von  $B$  (für  $i \in \{1, \dots, n\}$  und  $j \in \{1, \dots, r\}$ ), also

$$c_{i,j} = \sum_{\ell=1}^m a_{i,\ell} \cdot b_{\ell,j}.$$

Wieso reden wir plötzlich über das Vektor-Matrix Produkt?

Sei  $(G, P)$  eine Markov-Kette. Wenn wir mit Wahrscheinlichkeit  $\pi_i$  im Knoten  $i$  starten, dann sind wir nach einem Schritt im Knoten  $j$  mit Wahrscheinlichkeit

$$\sum_{i=1}^n \pi_i \cdot P_{i,j} \stackrel{\text{toll!}}{=} (\pi \cdot P)_j.$$

(Wir interpretieren Verteilungen als Zeilenvektoren.) Die Kette,

wenn in Verteilung  $\pi$  gestartet,

**befindet sich** nach einem Schritt **in der Verteilung**

$$\pi \cdot P,$$

denn  $(\pi \cdot P)_j$  ist die Wahrscheinlichkeit, den Zustand  $j$  in einem Schritt zu erreichen.



Und wenn wir die Markov-Kette  $k$  Schritte lang beobachten?

(a) **Rekursionsanfang** für  $k = 0$ : Die Kette befindet sich in der Verteilung

$$\pi^{(0)} := \pi.$$

(b) **Rekursionsschritt**: Wenn sich die Kette nach  $k$  Schritten in der Verteilung  $\pi^{(k)}$  befindet, dann befindet sie sich nach  $k + 1$  Schritten in der Verteilung

$$\pi^{(k+1)} := \pi^{(k)} \cdot P.$$

Mit vollständiger Induktion nach  $k$ :

Nach  $k$  Schritten befindet sich die Kette in der Verteilung

$$\pi^{(k)} = \pi \cdot P^k,$$

wenn wir in der Verteilung  $\pi$  beginnen.

Was ist die „**Grenzverteilung**“? Existiert der Grenzwert  $\lim_{k \rightarrow \infty} \pi \cdot P^k$ ?

# Die Berechnung von Matrizenprodukten

Wie schnell – wenn überhaupt – konvergieren die Verteilungen

$$\pi^{(k)} = \pi \cdot P^k$$

Werkzeuge, die das Matrizenprodukt  $P^k$  und das Vektor-Matrixprodukt  $\pi \cdot P^k$  schnell (und komfortabel) berechnen:

1. Für kleine Markov-Ketten ist das in

<https://matrixcalc.org/de/>

bereitgestellte Online-Werkzeug völlig ausreichend.

2. Für größere Ketten ist SymPy

<http://docs.sympy.org/latest/tutorial/matrices.html>

eine gute Wahl.

# Zusammenfassung: Was ist bisher passiert?

(a) Wie definiert man das **Renommee einer Webseite**  $w$ ?

- ▶ Mit  $PR_w$ , falls das Tupel PR die **Page-Rank-Eigenschaft** besitzt oder
- ▶ mit der relativen Häufigkeit  $PR_w^*$  mit der der **Zufallssurfer** die Seite  $w$  besucht?

(b) Der Zufallssurfer unternimmt eine **Irrfahrt** auf  $G$ . Mit einer Markov-Kette  $\mathcal{M} = (G, P)$  können wir Irrfahrten im gerichteten Graphen  $G$  untersuchen.

- ▶ Wenn  $(i, j)$  eine Kante von  $G$  ist, dann läuft die Irrfahrt mit Wahrscheinlichkeit

$$P_{i,j}$$

in einem Schritt von Zustand  $i$  nach Zustand  $j$

(c) Irrfahrten und das **Vektor-Matrix-Produkt**:

Wenn sich die Kette  $\mathcal{M} = (G, P)$  anfänglich in der Verteilung  $\pi$  befindet,

- ▶ (d.h. Zustand  $i$  wird mit Wahrscheinlichkeit  $\pi_i$  besucht)

dann befindet sie sich nach  $k$  Schritten im Zustand  $\pi \cdot P^k$ .

- ▶ (d.h. Zustand  $i$  wird mit Wahrscheinlichkeit  $(\pi \cdot P^k)_i$  besucht)

Markov-Ketten: Was möchte man gerne wissen?

# Markov-Ketten: Alter und neuer Page-Rank

Der vollständige, gerichtete Graph  $\vec{K}_n = (\mathbf{V}, \mathbf{E}_n)$  besitzt Knotenmenge  $V = \{1, \dots, n\}$  und Kantenmenge  $E_n = \{(u, v) : u, v \in \{1, \dots, n\}\}$ .

Sei  $\text{WEB}$  der Webgraph. Die **Webkette**  $\mathcal{W}$  wird beschrieben durch das Paar

$$\mathcal{W} := (\vec{K}_n, P_d(\text{WEB})).$$

Die Webkette muss uns helfen, den alten und neuen Page-Rank zu verstehen!

Was möchte man gerne wissen?

- ? Wie hängen PR und PR\*, also alter und neuer Page-Rank zusammen?
  - ▶ Existiert der alte Page-Rank überhaupt?
  - ▶ Hängt der neue Page-Rank von der Startverteilung des Zufallssurfers ab?
- ? Lässt sich der alte, bzw. der neue Page-Rank effizient berechnen?

# Markov-Ketten: Irrfahrten auf ungerichteten Graphen (1/2)

Viele Mähroboter, die ohne GPS-Ortung arbeiten, mähen den Rasen nach dem Zufallsprinzip.

- Der Roboter fährt gerade Strecken über den Rasen und wendet dann am Rand zufällig in einem von endlich vielen Winkeln.
- Wird tatsächlich jede Stelle des Rasens hochwahrscheinlich gemäht?

- (a) Sei  $G = (V, E)$  ein ungerichteter, zusammenhängender Graph  $G = (V, E)$ .
- ▶ Sei  $d_v$  die Anzahl der Nachbarn von  $v$ .
  - ▶ Hat die Irrfahrt den Knoten  $v$  erreicht, dann wird die Irrfahrt mit Wahrscheinlichkeit  $1/d_v$  mit einem Nachbarn  $v_i$  ( $1 \leq i \leq d_v$ ) von  $v$  fortgesetzt.
- (b) Die Markov-Kette  $(G', P)$  dieser Irrfahrt besitzt den Graphen  $G' = (V, E')$ ,
- ▶ wobei  $G$  und  $G'$  dieselbe Knotenmenge  $V$  besitzen
  - ▶ und eine **gerichtete** Kante  $(i, j)$  genau dann in  $G'$  vorkommt, wenn  $\{i, j\}$  eine (**ungerichtete**) Kante von  $G$  ist,
- sowie die Übergangsmatrix  $P$  mit

$$P_{i,j} := \begin{cases} \frac{1}{d_i} & \text{falls } \{i, j\} \text{ eine Kante von } G \text{ ist,} \\ 0 & \text{sonst.} \end{cases}$$

## Was möchte man gerne wissen?

- (?) Ab welcher Länge besucht eine Irrfahrt wahrscheinlich alle Knoten von  $G$ ?
- ▶ Die erwartete Länge einer Irrfahrt, die alle Knoten des Graphen  $G = (V, E)$  besucht, ist höchstens  $2|V| \cdot |E|$ . (Siehe Vorlesung „Effiziente Algorithmen“.)  
*Man sagt auch, dass die „Cover-Time“ des Graphen durch  $2|V| \cdot |E|$  beschränkt ist.*
  - ▶ Für reguläre Graphen, also für Graphen in denen jeder Knoten dieselbe Anzahl von Nachbarn besitzt, ist die Cover-Time höchstens  $4 \cdot |V|^2$ .
- (?) Was ist die relative Häufigkeit mit der ein Knoten  $i \in V$  in einer genügend langen Irrfahrt besucht wird?
- ▶ Antwort später.

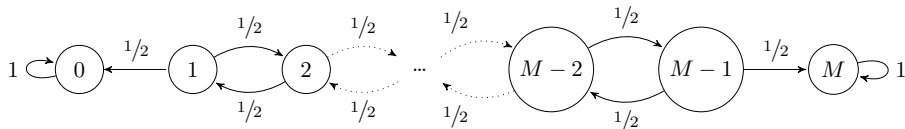
Ein Spieler tritt gegen das Casino an:

Bei einem Einsatz von 1 € ist der Gewinn/Verlust in jeder Runde ebenfalls 1 €.

Der Spieler hat ein Kapital von  $K$  €, das Casino von  $N$  €.

Die **Markov-Kette**:

- Wir verwenden die Zustände  $0, \dots, M$  mit  $M = K + N$
- und erlauben Übergänge vom Zustand  $0 < i < n$  zu den Zuständen  $i - 1$  und  $i + 1$  mit der Wahrscheinlichkeit von jeweils  $1/2$ .
  - Die Kette beginnt im Zustand  $K$ .
  - Im Zustand  $0$  ist der **Spieler ruiniert**, im Zustand  $M$  das **Casino gesprengt**:  
(Die Kette verbleibt im Zustand  $0$  bzw.  $M$  mit Wahrscheinlichkeit 1.)





## Was möchte man gerne wissen?

- ? Wie groß ist die Wahrscheinlichkeit  $s_K$ , dass der Spieler die Bank sprengt, wenn Casino und Spieler bis zum bitteren Ende spielen?
  - ▶ Es ist  $s_K = K/M$ .
  - ▶ Relativ gute Chancen, die Bank zu sprengen, selbst wenn  $K$  viel kleiner als  $M$  ist!
- ? Deshalb sind leider in professionellen Casinos alle Spiele unfair.
  - ▶ Sei  $p$  die Gewinnwahrscheinlichkeit des Spielers und es gelte  $p \neq 1/2$ . Weiterhin sei  $q := 1 - p$  die Komplementärwahrscheinlichkeit.
  - ▶ Es kann gezeigt werden:

$$s_K = \frac{\left(\frac{q}{p}\right)^K - 1}{\left(\frac{q}{p}\right)^M - 1}.$$

Für  $p < 1/2$  gilt

$$\frac{\left(\frac{q}{p}\right)^K - 1}{\left(\frac{q}{p}\right)^M - 1} \leq \frac{\left(\frac{q}{p}\right)^K}{\left(\frac{q}{p}\right)^M} = \left(\frac{q}{p}\right)^{-N}.$$

Die Erfolgswahrscheinlichkeit  $s_K$  für den Spieler fällt **exponentiell** mit  $N$  :-(((

Die KNF-Formel

$$\alpha = \bigwedge_{i=1}^m (\ell_{i,1} \vee \ell_{i,2})$$

besitzt höchstens 2 Literale pro Disjunktionsterm. Ist  $\alpha$  erfüllbar?

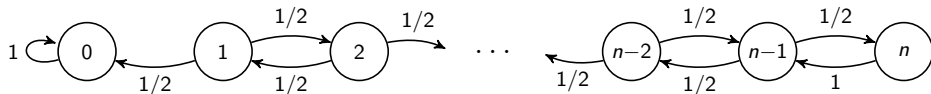
**Annahme:**  $\alpha$  besitze  $n$  Variablen und werde von der Belegung  $x^* \in \{0, 1\}^n$  erfüllt.

**Ein würfelnder Algorithmus:**

1. Würfle eine zufällige Belegung  $x \in \{0, 1\}^n$  aus.
2. Wiederhole **genügend oft**:
  - (a) Bestimme irgendeinen von  $x$  falsifizierten Disjunktionsterm  $D = \ell \vee \ell'$ .
  - (b) Wähle  $L \in \{\ell, \ell'\}$  zufällig aus: Modifiziere  $x$ , so dass  $L$  erfüllt wird, d.h. „flippe“ die Belegung der Variablen von  $L$ .  
 #  $x^*$  erfüllt  $D$  und deshalb auch  $\ell$  oder  $\ell'$ .  
 # Mit Wahrscheinlichkeit  $\geq \frac{1}{2}$  stimmen  $x, x^*$  in einer Variablen mehr überein,  
 # mit Wahrscheinlichkeit  $\leq \frac{1}{2}$  stimmen  $x, x^*$  in einer Variablen weniger überein.

Wie häufig muss wiederholt werden, bis eine erfüllende Belegung gefunden wird?

Wie verhält sich  $H(x, x^*)$ , der Hammingabstand<sup>1</sup> zwischen  $x$  und  $x^*$  im schlimmsten Fall? Analysiere die folgende Markov-Kette:



- Zustand  $i$  entspricht dem aktuellen Hammingabstand  $i$ . Insbesondere haben wir  $x^*$  im Zustand  $0$  gefunden.
- Zustandsübergänge protokollieren den Verlauf des Algorithmus.

Aufgabe 9.4: Der Algorithmus findet  $x^*$  nach einer erwarteten Anzahl von höchstens  $n^2$  Wiederholungen, wenn nicht vorher eine andere erfüllende Belegung gefunden wird.

<sup>1</sup> $H(x, x^*)$  ist die Anzahl der Variablen, für die sich  $x$  und  $x^*$  unterscheiden.

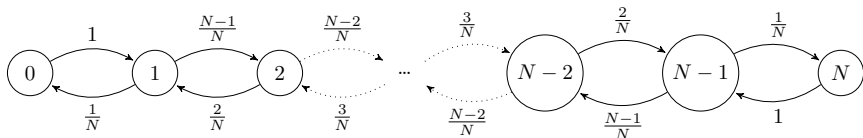
Zwei Substanzen sind durch eine Membran getrennt, aber Moleküle wandern über die Membran zwischen den beiden Substanzen.

**Modellierung:** In jedem Schritt wird eines der  $N$  Partikel gleichverteilt, also mit Wahrscheinlichkeit  $1/N$  ausgewählt und auf die jeweils andere Seite verschoben.

Unsere **Markov-Kette**  $(G, P)$  besitzt den Graphen  $G = (V, E)$  mit

- Knotenmenge  $V = \{0, 1, \dots, N\}$
- und den Kanten  $i \rightarrow i + 1$  und  $i \rightarrow i - 1$  für  $i = 1, \dots, N - 1$  sowie den Kanten  $0 \rightarrow 1, N \rightarrow N - 1$

und die Übergangsmatrix  $P$  mit  $P_{i,j} := \begin{cases} \frac{i}{N} & \text{falls } j = i - 1, \\ 1 - \frac{i}{N} & \text{falls } j = i + 1, \\ 0 & \text{sonst.} \end{cases}$



Was möchte man gerne wissen?

Angenommen, wir lassen Partikel „hinreichend oft“ wandern:

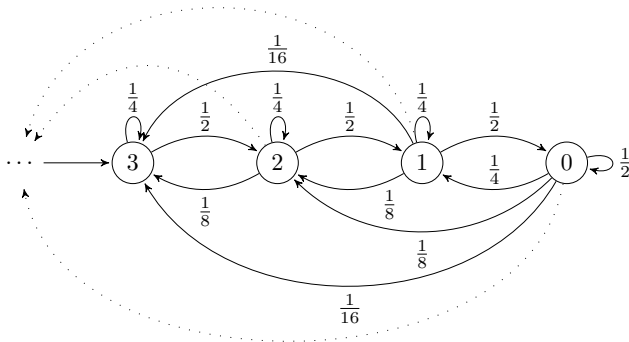
*Mit welcher Wahrscheinlichkeit besitzt die linke Seite genau  $l$  Partikel?*

Wir werden diese Frage später beantworten.

# Markov-Ketten: Warteschlangen

Wir modellieren die **Warteschlange** vor einer Supermarktkasse:

- Die **unendliche** Menge  $V = \mathbb{N}$  ist Zustandsmenge:  
Im Zustand  $i \in V$  warten  $i$  Kunden an der Kasse.
- Pro Zeitschritt bezahlt ein wartender Kunde, neue Kunden treffen ein.  
Z.B. treffen  $k$  neue Kunden mit Wahrscheinlichkeit  $2^{-k-1}$  ein.



**Was möchte man gerne wissen?** Was ist die erwartete Länge der Warteschlange?

# Zusammenfassung: Was ist bisher passiert?

- (a) Wie definiert man das **Renommee einer Webseite**  $w$ ?
- ▶ Mit  $PR_w$ , falls das Tupel PR die **Page-Rank-Eigenschaft** besitzt oder
  - ▶ mit der Wahrscheinlichkeit  $PR_w^*$  mit der der **Zufallssurfer** die Seite  $w$  besucht?
- (b) Der Zufallssurfer unternimmt eine **Irrfahrt** auf  $G$ . Mit einer Markov-Kette  $\mathcal{M} = (G, P)$  können wir Irrfahrten im gerichteten Graphen  $G$  untersuchen.
- ▶ Wenn  $(i, j)$  eine Kante von  $G$  ist, dann läuft die Irrfahrt mit Wahrscheinlichkeit

$$P_{i,j}$$

in einem Schritt von Zustand  $i$  nach Zustand  $j$

- (c) Irrfahrten und das **Vektor-Matrix-Produkt**:

Wenn sich die Kette  $\mathcal{M} = (G, P)$  anfänglich in der Verteilung  $\pi$  befindet,

- ▶ (d.h. Zustand  $i$  wird mit Wahrscheinlichkeit  $\pi_i$  besucht)
- dann befindet sie sich nach  $k$  Schritten im Zustand  $\pi \cdot P^k$ .
- ▶ (d.h. Zustand  $i$  wird mit Wahrscheinlichkeit  $(\pi \cdot P^k)_i$  besucht)

Gilt  $PR = PR^*$ ?

# Grenzverteilung und Grenzmatrix



# Die Grenzverteilung einer Markov-Kette

$\mathcal{M} = (G, P)$  sei eine Markov-Kette. Die Verteilung

$$\mathcal{G}(\mathcal{M})$$

heißt **Grenzverteilung der Kette**, wenn für **alle** Anfangsverteilungen  $\pi$  gilt:

$$\lim_{k \rightarrow \infty} \pi \cdot P^k = \mathcal{G}(\mathcal{M}).$$

Schließlich ist  $P^\infty := \lim_{k \rightarrow \infty} P^k$  die **Grenzmatrix**.

Sei  $\mathcal{W} := (\vec{K}_n, P_d(\text{WEB}))$  die **Webkette**.

**Wenn** die Webkette  $\mathcal{W}$  eine Grenzverteilung und Grenzmatrix besitzt, dann ist

$$\text{PR}^* = \mathcal{G}(\mathcal{W}) = \lim_{k \rightarrow \infty} \pi \cdot (P_d(\text{WEB}))^k \stackrel{!}{=} \pi \cdot \lim_{k \rightarrow \infty} P^k = \pi \cdot P^\infty.$$

Existiert die Grenzverteilung überhaupt?

Wenn  $\mathcal{M} = (G, P)$  eine schöne Markov-Kette ist,

(a) dann existiert die Grenzmatrix

$$P^\infty := \lim_{k \rightarrow \infty} P^k$$

(b) und die Irrfahrt **vergisst** ihren Startpunkt: Für alle  $i, i^*, j$  gilt

$$(P^\infty)_{i,j} = \lim_{k \rightarrow \infty} (P^k)_{i,j} = \lim_{k \rightarrow \infty} (P^k)_{i^*,j} = (P^\infty)_{i^*,j}.$$

Also sind alle Zeilen von  $P^\infty$  identisch.

Wenn  $\mathcal{M}$  schön ist, dann existieren Grenzwerte und alle Zeilen der Grenzmatrix  $P^\infty$  stimmen überein mit der ersten Zeile

$$z = ((P^\infty)_{1,j} : 1 \leq j \leq n) = ((\lim_{k \rightarrow \infty} P^k)_{1,j} : 1 \leq j \leq n)$$

$$\implies \text{Für alle } \pi \text{ gilt: } \pi \cdot P^\infty = \pi \cdot \begin{pmatrix} z \\ \vdots \\ z \end{pmatrix} = z$$

# Ergodische Ketten: Schöne Markov-Ketten

# „schön“ = ergodisch

Eine Markov-Kette  $(G, P)$  mit  $G = (V, E)$  ist **ergodisch**, wenn

- (a) die Grenzwahrscheinlichkeiten  $\lim_{k \rightarrow \infty} (P^k)_{i,j}$  und  $\lim_{k \rightarrow \infty} (P^k)_{i^*,j}$  für alle Zustände  $i, i^*, j$  existieren und übereinstimmen sowie
- (b)  $\lim_{k \rightarrow \infty} (P^k)_{i,j} > 0$  für alle Zustände  $i, j$  gilt.

Die Markov-Kette  $\mathcal{M} = (G, P)$  sei ergodisch und  $z$  sei die erste Zeile von  $P^\infty$ . Dann gilt für jede Verteilung  $\pi$

$$\pi \cdot P^\infty = z = \left( \lim_{k \rightarrow \infty} (P^k)_{1,1}, \dots, \lim_{k \rightarrow \infty} (P^k)_{1,n} \right) = \mathcal{G}(\mathcal{M}).$$

**Wenn** die Webkette  $\mathcal{W}$  ergodisch ist, ist

$$PR^* = \pi \cdot P^\infty = \mathcal{G}(\mathcal{W})$$

für jede Anfangsverteilung  $\pi$  und  $PR^*$  stimmt mit jeder Zeile von  $P^\infty$  überein.

# Das Leben ist hart, oder?

Der Graph  $G$  sei **bipartit**.

- Wenn wir eine Irrfahrt in einem Knoten  $i$  beginnen, ist

$$(P^{2k+1})_{i,i} = 0,$$

denn ein bipartiter Graph hat keine Kreise ungerader Länge.

- Der Grenzwert  $\lim_{k \rightarrow \infty} (P^k)_{i,i}$  existiert nicht, wenn  $\lim_{k \rightarrow \infty} (P^{2k})_{i,i} > 0$  gilt

:-((

Aber der vollständige Graph  $\vec{K}_n$  der Webkette ist doch nicht bipartit!

:-))

# Es ist bald Weihnachten!

Was sollten wir uns von dem Graphen  $G$  einer schönen, d.h. ergodischen Kette  $\mathcal{M} = (G, P)$  wünschen?

Sei  $G$  ein gerichteter Graph.

- (a)  $G$  heißt genau dann **irreduzibel**, wenn  $G$  stark zusammenhängend ist.
- (b) Ein Zustand  $i \in V$  hat die **Periode**  $p$ , wenn die Längen *aller* (nicht notwendigerweise einfachen) Wege von  $i$  nach  $i$  durch  $p$  teilbar sind.  
 $G$  heißt **aperiodisch**  $\iff$  **kein** Zustand besitzt eine Periode  $p > 1$ .

Man kann zeigen:

Die Markov-Kette  $(G, P)$  ist ergodisch  $\iff G$  ist irreduzibel und aperiodisch.

$\implies$ :

- i)  $G$  muss irreduzibel sein, denn  $\lim_{k \rightarrow \infty} (P^k)_{i,j} > 0$  wird für alle  $i, j$  gefordert.
- ii) Der Zustand  $i$  darf keine Periode haben, denn sonst existiert  $\lim_{k \rightarrow \infty} (P^k)_{i,i}$  nicht!

$\impliedby$ : Siehe die „Stochastik für die Informatik“ (Stl).

# Was ist bisher passiert?

- (a) Wie definiert man das **Renommee einer Webseite**  $w$ ?
- ▶ Mit  $PR_w$ , falls das Tupel  $PR$  die **Page-Rank-Eigenschaft** besitzt oder
  - ▶ mit der Wahrscheinlichkeit  $PR_w^*$  mit der der **Zufallssurfer** die Seite  $w$  besucht?
- (b) Der Zufallssurfer unternimmt eine **Irrfahrt** auf  $G$ . Mit einer **Markov-Kette**

$$\mathcal{M} = (G, P)$$

können wir Irrfahrten im gerichteten Graphen  $G$  untersuchen.

- (c) Irrfahrten und das **Vektor-Matrix-Produkt**: Befindet sich  $\mathcal{M} = (G, P)$  in Verteilung  $\pi$ , dann ist sie nach  $k$  Schritten in Verteilung

$$\pi \cdot P^k.$$

- (d)  $\lim_{k \rightarrow \infty} P^k = P^\infty$  ist die **Grenzmatrix**.  $\mathcal{M}$  hat die **Grenzverteilung**  $\mathcal{G}(\mathcal{M})$ , wenn  $\mathcal{G}(\mathcal{M}) = \pi \cdot P^\infty$  f.a. Verteilungen  $\pi$  gilt.
- (e) Für eine **ergodische** Kette  $(G, P)$  – d.h.  $G$  ist **irreduzibel** und **aperiodisch** – stimmt jede Zeile von  $P^\infty$  mit der Grenzverteilung überein.

Gilt  $PR = PR^*$ ?



- (a) Die Webkette  $\mathcal{W} = (\vec{K}_n, P_d(\text{WEB}))$  ist **ergodisch**, denn für alle Knoten  $i, j$  besitzt  $\vec{K}_n$  eine Kante  $(i, j)$ .
- ▶  $\vec{K}_n$  ist sicherlich irreduzibel, denn es gibt eine Kante zwischen je zwei Knoten
  - ▶ und aperiodisch, denn wir können in **einem** Schritt von  $i$  nach  $i$  laufen.

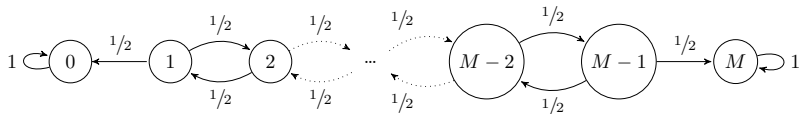
Aber, wie bestimmt man die Grenzverteilung  $\mathcal{G}(\mathcal{W})$ ?

- (b) Die Markov-Kette  $\mathcal{M} = (G', P)$  zu einer Irrfahrt auf einem **zusammenhängenden, nicht-bipartiten** Graphen  $G = (V, E)$  ist **ergodisch**.

- ▶ Zur Erinnerung: Es ist  $G' = (V, \{(i, j) : \{i, j\} \in E\})$ .
- ▶  $G$  ist zusammenhängend  $\implies$  der Graph  $G'$  der Kette ist irreduzibel.
- ▶ Warum ist  $G'$  aperiodisch? Sei  $u$  ein Zustand von  $G'$ .
  - ★  $u$  ist Teil eines Kreises in  $G'$  der Länge 2: Laufe von  $u$  zu einem Nachbarn und zurück.
  - ★  $G$  ist genau dann nicht bipartit, wenn  $G$  einen Kreis ungerader Länge hat.

Aber, wie bestimmt man die Grenzverteilung  $\mathcal{G}(\mathcal{M})$ ?

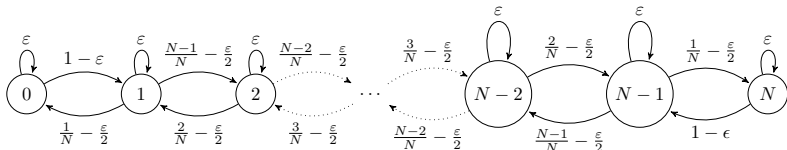
(c) „Gambler's Ruin“: Die Kette



ist **nicht ergodisch**, denn die Zustände 0 und  $M = K + N$  sind „**absorbierend**“ und der Graph ist nicht irreduzibel.

(d) Die Ehrenfest-Kette ist **nicht ergodisch**: Der Graph der Kette ist bipartit.

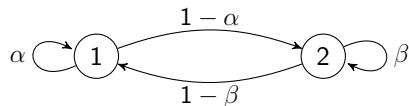
- ▶ Führe Eigenschleifen mit Wahrscheinlichkeit  $\epsilon$  ein.



- ▶ Jeder Knoten hat die Periode 1 aufgrund der Eigenschleifen: Die neue Kette ist also irreduzibel wie auch aperiodisch und deshalb **ergodisch**.
- ▶ Wie sieht ihre Grenzverteilung aus?

(e) Die „Warteschlangenkette“ ist irreduzibel und aperiodisch, aber sie besitzt **unendlich viele Zustände**: Im Allgemeinen folgt Ergodizität leider nicht.

(f) Für welche Werte von  $\alpha, \beta$  ist die folgende Kette ergodisch?



# Der alte und der neue Page-Rank

- ! Der neue Page-Rank  $PR^*$  stimmt mit der Grenzverteilung  $\mathcal{G}(W)$  der Webkette überein.
- ? Stehen alter und neuer Page-Rank miteinander in Beziehung?
- ? Lässt sich der neue, bzw. der alte Page-Rank effizient berechnen?

# Stationäre Verteilungen

# Stationäre Verteilungen

Sei  $(G, P)$  eine Markov-Kette mit  $G = (V, E)$  und es gelte  $V = \{1, \dots, n\}$ .

Eine Verteilung  $\pi$  auf  $V$  heißt **stationär** für die Markov-Kette  $(G, P)$ , falls

$$\pi \cdot P = \pi.$$

- Die Kette  $(G, P)$ , wenn in der Verteilung  $\pi$  gestartet, befindet sich nach einem Schritt in der Verteilung  $\pi \cdot P = \pi$ .
- Die Verteilung  $\pi$  ist stationär für  $(G, P) \iff$  die Kette bleibt in  $\pi$  „stecken“.

Der Page-Rank PR ist eine Verteilung, denn

$$\begin{aligned}
 \sum_{j=1}^n \text{PR}_j &\stackrel{\text{Definition}}{=} \sum_{j=1}^n \left( \frac{1-d}{n} + d \cdot \sum_{i \in \text{Vor}(j)} \frac{\text{PR}_i}{a_i} \right) = n \cdot \frac{1-d}{n} + d \cdot \sum_{j=1}^n \sum_{i \in \text{Vor}(j)} \frac{\text{PR}_i}{a_i} \\
 &= (1-d) + d \cdot \sum_{(i,j) \in E} \frac{\text{PR}_i}{a_i} = (1-d) + d \cdot \sum_{i=1}^n \sum_{j: (i,j) \in E} \frac{\text{PR}_i}{a_i} \\
 &= (1-d) + d \cdot \sum_{i=1}^n (a_i \cdot \frac{\text{PR}_i}{a_i}) = (1-d) + d \cdot \sum_{i=1}^n \text{PR}_i
 \end{aligned}$$

Also folgt  $(1-d) \cdot \sum_{j=1}^n \text{PR}_j = (1-d)$  bzw.  $\sum_{j=1}^n \text{PR}_j = 1$  ✓

Sei  $P_d(\text{WEB})$  die Page-Rank-Matrix, also

$$(P_d(\text{WEB}))_{i,j} = \frac{1-d}{n} + d \cdot \begin{cases} \frac{1}{a_i} & i \in \text{Vor}(j), \\ 0 & \text{sonst} \end{cases}$$

Der Page-Rank ist stationär, denn

$$\begin{aligned} \text{PR}_j & \stackrel{\text{Definition von PR}}{=} \frac{1-d}{n} \cdot 1 + d \cdot \sum_{i \in \text{Vor}(j)} \frac{\text{PR}_i}{a_i} \\ \text{PR ist eine Verteilung} & \quad \frac{1-d}{n} \cdot \sum_{i \in \{1, \dots, n\}} \text{PR}_i + d \cdot \sum_{i \in \text{Vor}(j)} \frac{\text{PR}_i}{a_i} \\ & = \sum_{i \in \{1, \dots, n\}} \text{PR}_i \cdot \left( \frac{1-d}{n} + d \cdot \begin{cases} \frac{1}{a_i} & i \in \text{Vor}(j), \\ 0 & \text{sonst} \end{cases} \right) \\ & \stackrel{\text{Definition von } P}{=} \sum_{i \in \{1, \dots, n\}} \text{PR}_i \cdot (P_d(\text{WEB}))_{i,j} = (\text{PR} \cdot P_d(\text{WEB}))_j \quad \checkmark \end{aligned}$$



Und jetzt Geschenke auspacken!

# Der Hauptsatz für ergodische Markov-Ketten

$\mathcal{M} = (G, P)$  sei eine **ergodische** Markov-Kette.

Dann gibt es genau eine **stationäre Verteilung**  $\sigma$  und  $\sigma$  stimmt mit der **Grenzverteilung** überein, d.h. es gilt

$$\mathcal{G}(\mathcal{M}) = \sigma.$$

Eine wichtige Konsequenz: Für die Webkette gilt

$$\mathbf{PR} = \mathbf{PR}^*.$$

Alter und neuer Page-Rank stimmen überein!

Die stationäre Verteilung ist gleichzeitig Grenzverteilung! Und warum stimmt das?

## Schritt 1: Die Grenzverteilung $\mathcal{G}(\mathcal{M})$ ist stationär!

$$\begin{aligned}(\mathcal{G}(\mathcal{M}) \cdot P)_j &= \sum_{i=1}^n \left( \lim_{k \rightarrow \infty} (P^k)_{1,i} \right) \cdot P_{i,j} = \lim_{k \rightarrow \infty} \sum_{i=1}^n (P^k)_{1,i} \cdot P_{i,j} \\ &= \lim_{k \rightarrow \infty} P_{1,j}^{k+1} = \lim_{k \rightarrow \infty} P_{1,j}^k \\ &= (\mathcal{G}(\mathcal{M}))_j.\end{aligned}$$

Es bleibt zu zeigen, dass  $\mathcal{G}(\mathcal{M})$  die einzige stationäre Verteilung ist.

## Schritt 2: Es gibt genau eine stationäre Verteilung

Sei  $\pi$  eine beliebige stationäre Verteilung für die Kette  $(G, P)$ .

1. Dann ist  $\pi = \pi \cdot P$  und deshalb gilt  $\pi = \pi \cdot P^k$  für jede natürliche Zahl  $k$ .  
Es folgt  $\pi = \lim_{k \rightarrow \infty} \pi \cdot P^k = \pi \cdot P^\infty$ .
2. Aber das ist doch die Grenzverteilung, denn

$$\pi = \pi \cdot P^\infty = \mathcal{G}(\mathcal{M}).$$

Es ist  $\pi = \mathcal{G}(\mathcal{M})$  und *die* stationäre Verteilung stimmt mit *der* Grenzverteilung überein.

# Stationäre Verteilungen: Beispiele

# Irrfahrten in ungerichteten Graphen

Sei  $G = (V, E)$  ein ungerichteter, zusammenhängender und nicht-bipartiter Graph. Dann besucht eine Irrfahrt den Knoten  $v$  mit Wahrscheinlichkeit

$$\pi_v = \frac{d_v}{2|E|}.$$

Ein Knoten wird mit Wahrscheinlichkeit proportional zu seinem Grad  $d_v$  besucht!

- ✓ Die Markov-Kette für Irrfahrten in ungerichteten, nicht-bipartiten Graphen ist ergodisch.
- ! Zeige für  $d_v = |\text{Nachbarn von } v|$ :  
Die Verteilung  $\pi$  mit  $\pi_v := \frac{d_v}{2|E|}$  ist stationär – sogar für bipartite Graphen.
- ! Zeige  $\pi \cdot P = \pi$  (für die Übergangsmatrix  $P$  eines beliebigen ungerichteten Graphen).

$$\begin{aligned}(\pi \cdot P)_v &= \sum_{u \in V} \pi_u \cdot P_{u,v} = \sum_{u \in V} \frac{d_u}{2|E|} \cdot P_{u,v} = \sum_{u \in V \text{ mit } \{u,v\} \in E} \frac{d_u}{2|E|} \frac{1}{d_u} \\ &= \sum_{u \in V \text{ mit } \{u,v\} \in E} \frac{1}{2|E|} = \frac{d_v}{2|E|} = \pi_v. \quad \checkmark\end{aligned}$$

# Symmetrische Ketten

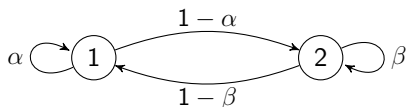
Sei  $(G, P)$  eine Markov-Kette und die Matrix  $P$  sei **symmetrisch**, d.h.  $P_{i,j} = P_{j,i}$  gilt für alle Knoten  $i, j$ . Dann ist die Gleichverteilung  $\sigma = (1/n, \dots, 1/n)$  stationär.

Warum?

$$\begin{aligned} (\sigma \cdot P)_j &= \sum_{i=1}^n \sigma_i \cdot P_{i,j} = \sum_{i=1}^n \frac{1}{n} \cdot P_{i,j} = \frac{1}{n} \cdot \sum_{i=1}^n P_{i,j} \\ &\stackrel{P \text{ ist symmetrisch}}{=} \frac{1}{n} \cdot \sum_{i=1}^n P_{j,i} \stackrel{P \text{ ist stochastisch}}{=} \frac{1}{n} = \sigma_j. \end{aligned}$$

Also gilt  $\sigma \cdot P = \sigma$  und die Gleichverteilung ist tatsächlich stationär  $\implies$  Jeder der  $n$  Zustände hat die Wahrscheinlichkeit  $\frac{1}{n}$ .

# Stationäre Verteilungen: Welches Gleichungssystem?



Eine Verteilung  $\sigma = (x, y)$  ist genau dann stationär für  $(G, P)$ , wenn  $\sigma \cdot P = \sigma$  gilt.

$$\begin{aligned}\sigma \cdot P &= (x, y) \cdot \begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \beta & \beta \end{pmatrix} \\ &= (x \cdot \alpha + y \cdot (1 - \beta), \quad x \cdot (1 - \alpha) + y \cdot \beta) \\ &\stackrel{!}{=} \sigma = (x, y).\end{aligned}$$

Also führt die Forderung  $\sigma P \stackrel{!}{=} \sigma$  auf das Gleichungssystem

$$\begin{aligned}x \cdot \alpha + y \cdot (1 - \beta) &\stackrel{!}{=} x \\ x \cdot (1 - \alpha) + y \cdot \beta &\stackrel{!}{=} y.\end{aligned}$$



Das Gleichungssystem

$$\begin{aligned}x \cdot \alpha + y \cdot (1 - \beta) &\stackrel{!}{=} x \\x \cdot (1 - \alpha) + y \cdot \beta &\stackrel{!}{=} y.\end{aligned}$$

ist äquivalent mit

$$x \cdot (1 - \alpha) \stackrel{!}{=} y \cdot (1 - \beta).$$

Wir erhalten einen 1-dimensionalen Lösungsraum: Was ist passiert?

Wenn  $\sigma \cdot P = \sigma$ , dann gilt auch  $\lambda\sigma \cdot P = \lambda\sigma$  für jedes Vielfache  $\lambda \in \mathbb{R}$ .

Wir müssen fordern, dass  $\sigma$  eine Verteilung ist, d.h. das  $x + y = 1$  gilt.

## Und die Lösung ist:

Wir erhalten das neue Gleichungssystem

$$\begin{aligned}x \cdot (1 - \alpha) &= y \cdot (1 - \beta) \\ y &= 1 - x\end{aligned}$$

mit der Lösung

$$x = \frac{1 - \beta}{2 - \alpha - \beta} \text{ und } y = \frac{1 - \alpha}{2 - \alpha - \beta}.$$

Die Kette besitzt eine eindeutige stationäre Verteilung, falls  $\alpha < 1$  oder  $\beta < 1$ .

Beachte, dass die Kette genau dann ergodisch ist, wenn

$$(\alpha > 0 \text{ oder } \beta > 0) \text{ und } \alpha < 1 \text{ und } \beta < 1$$

Also besitzen auch einige nicht-ergodische Ketten eindeutige stationäre Verteilungen.

# Stationäre Verteilungen der (ergodischen) Ehrenfest-Kette

Die Verteilung  $\pi$  mit

$$\pi_i = \binom{n}{i} / 2^n$$

für  $i = 0, \dots, n$  ist stationär.

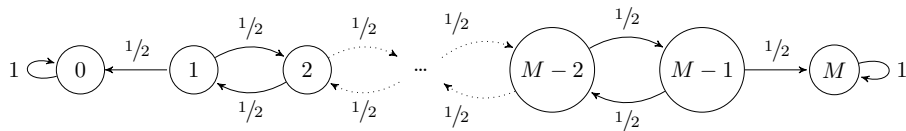
1.  $\pi$  ist die einzige stationäre Verteilung und
2.  $\pi$  stimmt mit der Grenzverteilung überein.

Auf der linken Seite der Membran befinden sich  $\ell$  Partikel mit Wahrscheinlichkeit

$$\binom{n}{\ell} / 2^n.$$

Man sagt, dass  $\pi$  eine **Binomialverteilung** ist.

# Stationäre Verteilungen im „Gambler's Ruin“



1. Wenn wir im Zustand 0 starten, bleiben wir stecken:  
Die Verteilung  $(1, 0, \dots, 0)$  ist stationär.
2. Gleiches gilt für den Zustand  $M$ : Die Verteilung  $(0, \dots, 0, 1)$  ist stationär.
3. Alle „Konvexkombinationen“, also alle Verteilungen

$$(\lambda, 0, \dots, 0, 1 - \lambda)$$

für  $0 \leq \lambda \leq 1$  sind stationär.

Diese Markov-Kette hat **unendlich** viele stationäre Verteilungen!

Lässt sich der Page-Rank schnell **und** gut approximieren?

Sei  $\mathcal{M} = (G, P)$  eine ergodische Markov-Kette.

1. Sei  $\pi^{(0)} := \pi$  eine beliebige Verteilung auf  $\{1, \dots, n\}$  und setze

$$\pi^{(k+1)} = \pi^{(k)} \cdot P.$$

2. Nach  $k$  Schritten befindet sich die Kette in der Verteilung

$$\pi^{(k)} := \pi \cdot P^k.$$

3. Wenn  $k$  genügend groß ist, dann ist

$$\pi^{(k)} = \pi \cdot P^k \approx \pi \cdot P^\infty = \mathcal{M}(\mathcal{G}).$$

- Es ist  $\pi^{(k)} \approx \mathcal{G}(\mathcal{M})$  wenn  $k$  „genügend“ groß ist.
- Um  $\pi^{(k)}$  zu bestimmen, berechne

$$\pi^{(1)} = \pi^{(0)} \cdot P, \pi^{(2)} = \pi^{(1)} \cdot P, \dots, \pi^{(k)} = \pi^{(k-1)} \cdot P.$$

- ✓ Die Berechnung eines jeden der  $k$  Vektor-Matrix Produkte ist „hochgradig parallelisierbar“.
- ✓  $\pi^{(k)} \approx \mathcal{G}(\mathcal{M})$  bereits für kleine Werte von  $k$ :
  - ▶ Das Web ist „hoch-gradig zusammenhängend“.
  - ▶ Die Hypothese des „**Kleine-Welt-Phänomens**“ besagt, dass in vielen sozialen Netzwerken (fast) jeder Mensch mit jedem anderen über eine Kette von höchstens **6** Bekanntschaftsbeziehungen verbunden ist.

:-))

Es ist alles gut!

:-))

# Zusammenfassung



# Zusammenfassung: Markov-Ketten

Die Kette  $\mathcal{M} = (G, P)$  besteht aus einer stochastischen Übergangsmatrix  $P$  und einem gerichteten Graphen  $G = (\{1, \dots, n\}, E)$  mit  $(i, j) \in E \iff P_{i,j} > 0$ .

- (a) Ein Schritt der Markov-Kette  $\mathcal{M}$  kann durch das Matrix-Vektor Produkt  $\pi \cdot P$  beschrieben werden:
- ▶ Wenn ein Zustand  $i$  mit Wahrscheinlichkeit  $\pi_i$  ausgewürfelt wird, dann befindet sich die Kette nach einem Schritt im Zustand  $j$  mit Wahrscheinlichkeit  $(\pi \cdot P)_j$ .
- (b) Eine **stationäre Verteilung**  $\sigma$  ist Lösung des linearen Gleichungssystems

$$\sigma \cdot P = \sigma.$$

- (c)  $(G, P)$  ist **ergodisch**, wenn  $G$  irreduzibel (bzw. stark zusammenhängend) und aperiodisch ist.
- ▶ **Der Hauptsatz für ergodische Markov-Ketten:** Eine ergodische Kette besitzt genau eine **stationäre Verteilung**  $\mathcal{G}(\mathcal{M})$  und  $\mathcal{G}(\mathcal{M})$  ist die **Grenzverteilung**.

# Zusammenfassung: Page-Rank

- (a) Durch das Hinzufügen von „neuen Kanten“ – entsprechend dem „Grundeinkommen“ – wird die Webkette **ergodisch**.
- (b) Der Ansatz des **Peer-Reviews**  $PR_v$ :
- ▶ Die Webseite  $v$  vererbt ihr Renommee anteilig auf jede Webseite  $w$ , auf die sie einen Hyperlink gesetzt hat.
  - ▶ Der Page-Rank ist die stationäre Verteilung der Webkette.
- (c) Der Ansatz des **Zufallssurfers**: Bestimme die relative Häufigkeit  $PR_v^*$  des Besuchs von  $v$ :
- ▶ Die Webkette ist ergodisch: Die Grenzwahrscheinlichkeit  $\pi_v$ , also die Wahrscheinlichkeit, dass der Zufallssurfer die Seite  $v$  am Ende einer langen Irrfahrt besucht, existiert.
  - ▶ Das Kleine-Welt-Phänomen: Die Konvergenz gegen  $\pi_v$  ist schnell.
- (d) Mit dem Hauptsatz für Markov-Ketten:  $PR = PR^*$
- ▶ Für vernünftige Anfangsverteilungen  $\pi$  und „kleine“ Werte von  $k$  gilt

$$\pi \cdot (P_d(\text{WEB}))^k \approx PR^* = PR.$$