POLITECNICO DI MILANO

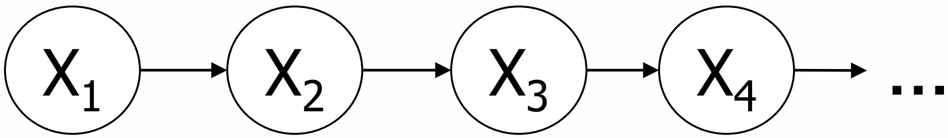
Markov Chains

Information Retrieval and Data Mining

Prof. Matteo Matteucci

Probabilistic Reasoning for Time Series

To describe an ever changing world we can use a series of random variables describing the world state at any time instant!



- It represents a sequence of states: X_1, X_2, X_3, …
- The transition from X_{t-1} to X_t depends only on X_{t-1}

$$P(X_t|X_{t-1}, X_{t-2}, \ldots, X_1, X_0) = P(X_t|X_{t-1})$$ (Markov Property)

- When transition probabilities are the same a any *t*, we are facing a stationary process.
- A *Bayesian Network* that forms a chain!

(Let's skip basic stuff and go to hidden models)

Stochastic Processes and Markov Chains

- Given X_t the value of a system characteristic at time *t* described as a (state) random variable, we have:
  - *Discrete Stochastic Process:* describes the a relationship between the stochastic description of a system (X_0, X_1, X_2, …) at some discrete time steps.
  - A *Continuous Stochastic Process* is a stochastic process where the state can be observed at any time.

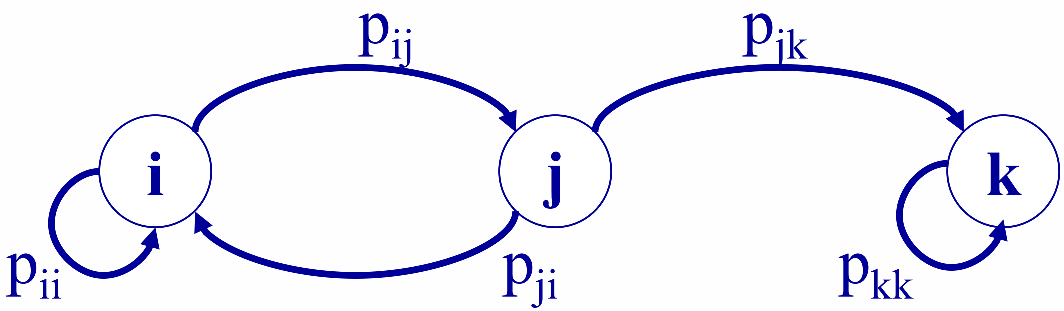
- A Discrete Stochastic Process is a (first order) ***Markov Chain*** when we have thet $\forall t = 1, 2, 3, \ldots$ and for all *n* states it holds:
  - $P(X_{t+1}=i_{t+1}|X_t=i_t, X_{t-1}=i_{t-1},\ldots,X_1=i_1,X_0=i_0)=P(X_{t+1}=i_{t+1}|X_t=i_t)$
- Whenever the probability of an event is independent from time the Markov Chain is *Stationary:* $P(X_{t+1}=j|X_t=i)=p_{ij}$

Markov Chain Description

- A Markov Chain can be described using a *Transition Matrix* where p_{ij} describes the probability of getting into state *j* starting from state *i*:

$$P = \begin{pmatrix} p_{11} p_{12} p_{13} \cdots p_{1n} \\ p_{21} p_{22} p_{23} \cdots p_{2n} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ p_{n1} p_{n2} p_{n3} \cdots p_{nn} \end{pmatrix} \qquad \sum_{j=1}^{n} p_{ij} = 1$$

- This transition matrix can be described also using a directed graph



Computing Probabilities

- Given a Markov Chain in state *i* at time *m* we can compute states probability after *n* time steps:

$$P(X_{m+n}=j|X_m=i)=P(X_n=j|X_0=i)=P_{ij}(n)$$

- If we take n=2 we have

$P_{ij}(2) = \sum_k p_{ik} \cdot p_{kj}$ Scalar product of row *i* and column *j*

- In general $P_{ij}(n)$ = *ij-th* element of P^{n}.

- The probability of being in a given state *j* at time *n* without knowing the exact state of Markov Chain at time *0* is thus:

$$\sum_i q_i \cdot P_{ij}(n) = q \cdot (\text{column } j \text{ of } P^n)$$

- where:

$$q_i = \text{state } i \text{ probability at time } 0$$

The Cola Example (I)

- Suppose our company produces two brands of Cola (i.e., Cola1, and Cola2) and there are no other Colas on the market. A person buying Cola1 will buy Cola1 again with probability 0.9. A persona buying Cola2 will buy Cola2 again with probability 0.8.

$$P = \begin{array}{c} \\ \text{Cola1} \\ \text{Cola2} \end{array} \begin{array}{c} \begin{array}{cc} \text{Cola1} & \text{Cola2} \end{array} \\ \begin{pmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{pmatrix} \end{array}$$

  - Someone has bought Cola2, what's the probability he/she will buy Cola1 after 2 times?
  - Someone has bought Cola1, what's the probability he/she will buy Cola1 again after 3 times?
  - Suppose at some time 60% of clients bought Cola1 and 40% Cola2. After three purchases what's the percentage of people buying Cola1?

The Cola Example (II)

- Someone has bought Cola2, what's the probability he/she will buy Cola1 after 2 times?
  - $P(X_2=1|X_0=2)=P_{21}(2)$

$$P2=\begin{pmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{pmatrix}\begin{pmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{pmatrix}=\begin{pmatrix} 0.83 & 0.17 \\ \quad & 0.66 \end{pmatrix}$$

- Someone has bought Cola1, what's the probability he/she will buy Cola1 again after 3 times?
  - $P(X_3=1|X_0=1)=P_{11}(3)$

$$P3=\begin{pmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{pmatrix}\begin{pmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{pmatrix}=\begin{pmatrix} \quad & 0.219 \\ 0.438 & 0.562 \end{pmatrix}$$

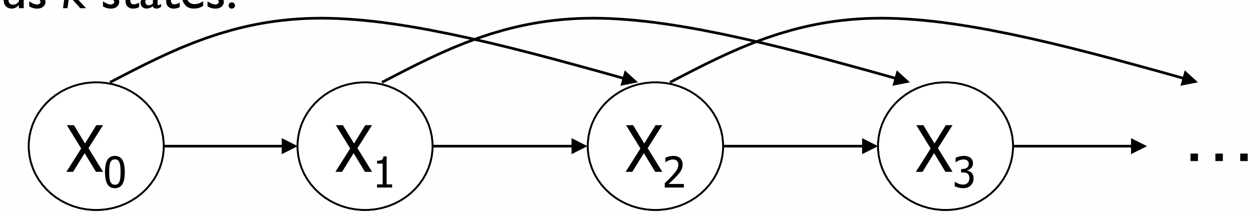
The Cola Example (III)

- Suppose at some time 60% of clients bought Cola1 and 40% Cola2. After three purchases what's the percentage of people buying Cola1?

$$p=\sum_i q_i \cdot P_{ij}(3) = q \cdot (\text{column } 1 \text{ of } P^3)$$

$$p= \begin{bmatrix} 0.60 & 0.40 \end{bmatrix} \begin{bmatrix} 0.781 \\ 0.438 \end{bmatrix} = 0.6438$$

- **Note:** What we have discussed so far is the first-order Markov Chain. More generally, in k^{th}-order Markov Chain, each state transition depends on previous k states.



What's the size of transition probability matrix?

A Bunch of Definitions

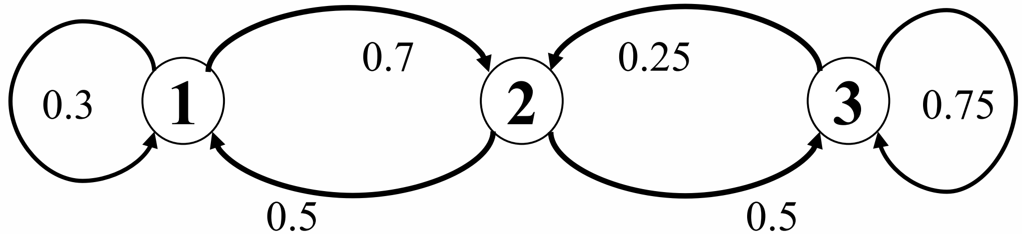
- Given a Markov Chain we define:
  - State *j* is ***reachable*** from *i* if it exist a path from *i* to *j*
  - States *i* and *j* ***communicate*** if *i* is reachable from *j* and viceversa
  - A set of states *S* in a Markov Chain is ***closed*** if no state outside S is reachable from a state in *S*
  - A state *i* is an ***absorbing*** state if $p_{ii}=1$
  - A state *i* is ***transient*** if exists *j* reachable from *i*, but *i* is not reachable from *j*
  - A state that is not transient is defined as ***recurrent***
  - A state *i* is ***periodic*** with period $k>1$ if *k* is the biggest number that divides the length of all path from *i* to *i*
  - A state that is not periodic is said ***a-periodic***

- If all states in a Markov Chain are *recurrent*, *a-periodic*, and *communicate* with each other, it is said to be **Ergothic**

Examples of Ergothic Markov Chains

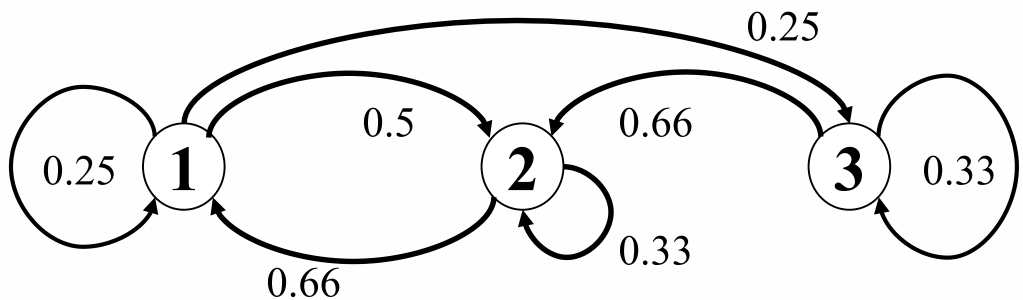
- A simple example of Ergothic Markov Chain is the following:

$$P = \begin{pmatrix} 0.3 & 0.7 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.25 & 0.75 \end{pmatrix}$$

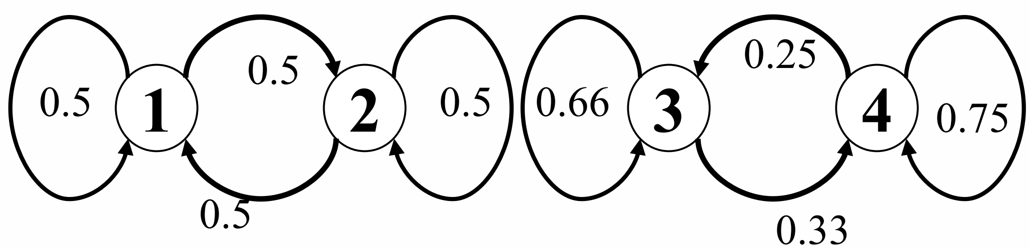


- Do the following transitions represent Ergothic Markov Chains?

$$P = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 2/3 & 1/3 & 0 \\ 0 & 2/3 & 1/3 \end{pmatrix}$$



$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 2/3 & 1/3 \\ 0 & 0 & 1/4 & 3/4 \end{pmatrix}$$



Steady State Distribution

- Being *P* the transition matrix of an Ergothic Markov Chain with *n* states we have that

$$\lim_{n \to +\infty} \boldsymbol{P}_{ij}(n) = \pi_j$$

- With $\pi = [\pi_1\ \pi_2\ \pi_3\ \dots\ \pi_n] = \pi \cdot P$ being the *Steady State Distribution*

- The Cola Example:

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$

$$\pi = \begin{pmatrix} 0.67 & 0.33 \\ 0.67 & 0.33 \end{pmatrix}$$

STEADY STATE

n	$P_{11}(n)$	$P_{12}(n)$	$P_{21}(n)$	$P_{22}(n)$
1	.90	.10	.20	.80
2	.83	.17	.34	.66
3	.78	.22	.44	.56
5	.72	.28	.56	.44
10	.68	.32	.65	.35
20	.67	.33	.67	.33
30	.67	.33	.67	.33
40	.67	.33	.67	.33

Transitory Behavior

- The behavior of a Markov Chain before getting to the Steady State id defined *transitory*

- We can compute the expected *number of transition* to reach state *j* being in state *i* for an Ergothic Markov Chain:

$$m_{ij} = p_{ij}(1) + \sum_{k \neq j} p_{ik} \cdot (1 + m_{kj}) = 1 + \sum_{k \neq j} p_{ik} \cdot m_{kj}$$

- The Cola Example:
  - How many bottle on average a Cola1 buyer will have before switching to Cola2?

$$m_{12} = 1 + \sum_{k \neq j} p_{1k} \cdot m_{k2} = 1 + 0.9\, m_{12} \quad \longrightarrow \quad m_{12} = 10$$

  - What about viceversa?

$$m_{21} = 1 + \sum_{k \neq j} p_{2k} \cdot m_{k1} = 1 + 0.8\, m_{21} \quad \longrightarrow \quad m_{21} = 5$$

Dealing with Absorbing States

- We have and *absorbing Markov Chain* if there exist one or more absorbing states and all the other are transient.
- For an absorbing Markov Chain we can write the transition matrix as:

$$P = \left(\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array} \right)$$

- where:
  - *Q* is the transition matrix for transient states
  - *R* is the trantion matrix from transient to absorbing states
- What kind of inference we could make with this model?
  - How long it will take to get in an absorbing state given that we start from a transient one?
  - Starting from a transient state, how long does it takes to get to an absorbing one?

Inference in Absorbing Markov Chains

- How long I remain in a transient state given that we start from a transient one?
  - Being in a transient state *i* the average time spent in a transient state *j* is the *ij*-th element of $(I-Q)^{-1}$
- Starting from a transient state, how long does it takes to get to an absorbing one?
  - Being in transient state *i* the probability to get into an absorbing state *j* is the *ij*-th element of $(I-Q)^{-1} \cdot R$
- <u>Example</u>: in a company there are 3 levels: junior, senior, partner. You can leave the company as partner or not
  - How long does a junior remains in the company?
  - What's the probability for a junior to leave the company as partner?

$$P = \begin{array}{c} \begin{array}{ccccc} J & S & P & LN & LP \end{array} \\ \left(\begin{array}{ccc|cc} 0.80 & 0.15 & 0 & 0.05 & 0 \\ 0 & 0.70 & 0.20 & 0.10 & 0 \\ 0 & 0 & 0.95 & 0 & 0.05 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right) \end{array}$$

The Company Example

- How long does a junior remains in the company?

$$(I-Q)^{-1} = \begin{pmatrix} 5 & 2.5 & 10 \\ 0 & 3.3 & 13.3 \\ 0 & 0 & 20 \end{pmatrix}$$

  - He/she will stay as Junior: $m_{11} = 5$
  - He/she will stay as Senior: $m_{12} = 2.5$
  - He/She will stay as Partner: $m_{13} = 10$

  } <u>17.5 years!</u>

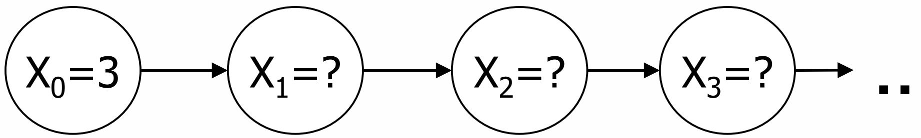
- What's the probability for a junior to leave the company as partner?

$$(I-Q)^{-1} \cdot R = \begin{pmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \\ 0 & 1 \end{pmatrix}$$

  - He/She will end up in state LP: $m_{12} = 0.5$

Exercise: Gambler's Ruin

- Suppose we are a gambler and we start from a 3$ capital, with probability $p=1/3$ we can win *1$* and with probability $1-p=2/3$ we loose *1$*. We fail if out capital get to 0 and we win if our capital becomes 5.



- We can describe our capital as a Markov Chain being X_t our capital:
  - Possible states: *0, 1, 2, 3, 4, 5*
  - Transition probability: $p(X_{t+1}=X_t+1)=1/3$, $p(X_{t+1}=X_t-1)=2/3$
- What kind of reasoning can we apply to this model?
  - What's the probability of sequence 3, 4, 3, 2, 3, 2, 1, 0?
  - What's the probability of success for the gambler?
  - What's the average number of bets the gambler will make?

Why Should I Care All This Crazy Math?

"Nice, but unless I want to gamble why should I care? I'm a computer engineer what this has to do with practical intelligent systems?"

- What do you this is the greatest revolution (or revolutionary company) on the web in the last decade?

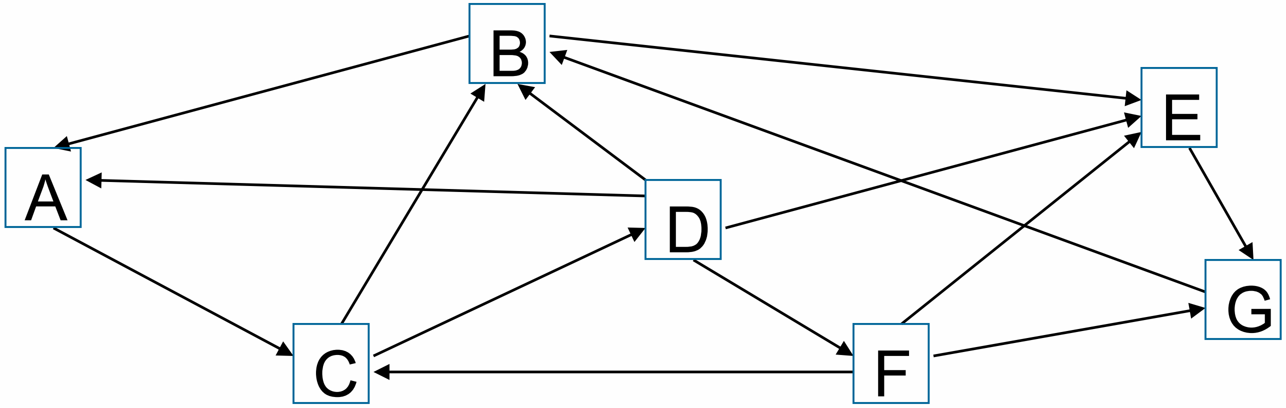
- Assume a link from page A to page B is a recommendation of page B by the author of A (we say B is successor of A).
  - Quality of a page is related to its *in-degree.*
  - The of a page is related to the quality of pages *linking to it*
- This recursively defines the ***PageRank*** of a page [Brin & Page '98]

For a (better) detailed description feel free to read:
http://www-db.stanford.edu/~backrub/google.html
http://www.iprcom.com/papers/pagerank/

Google's PageRank

- Suppose the web is an Ergothic Markov Chain (I know this is a big assumption). Consider browsing as an infinite random walk (surfing):
  - Initially the surfer is at a random page
  - At each step, the surfer proceeds
    - to a randomly chosen web page with probability d
    - to a randomly chosen successor of the current page with probability 1-d

- The PageRank of a page is the fraction of steps the surfer spends on it in the limit.

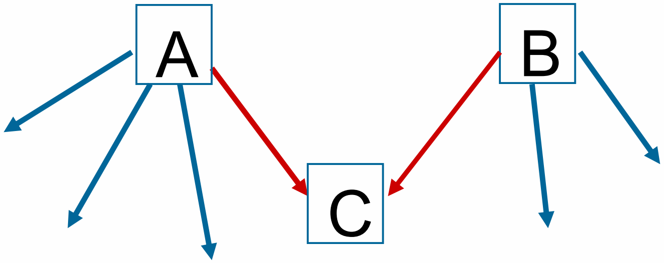


Definition of PageRank

- PageRank = the steady state probability for this Markov Chain

$$PageRank(u) = d + (1-d)\sum_{(v,u)\in E} PageRank(v)/outdegree(v)$$

  - *n* is the total number of nodes in the graph
  - *d* is the probability of a random jump



PageRank(C) = d/n+(1-d)(1/4 PageRan(A) +1/3 PageRank(B))

- Summarizes the "web opinion" about the page importance
  - Query-independent
  - It can be faked … read the provided links if you are curious!