CSE 312

Foundations of Computing II

Lecture 24: Markov Chains

So far: probability for "single-shot" processes

Random Process → **Outcome Distribution** D

Today:
A very special type of DTSP called **Markov Chains**

More generally: randomness can enter over many steps and depend on previous outcomes

x_1 $\quad$ $x_1 + x_2 + r$ $\quad$ x_3

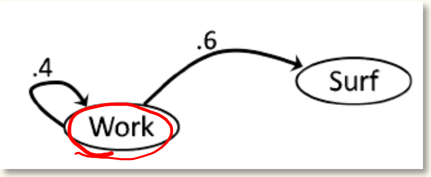
Random Process 1 → **Outcome Distribution** D_1 → Random Process 2 → **Outcome Distribution** D_2 → Random Process 3 → **Outcome Distribution** D_3 …

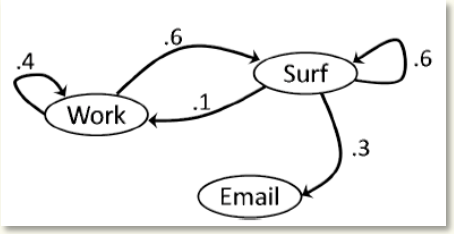
D_1 $\quad$ X_2 $\quad$ r $\quad$ D_2

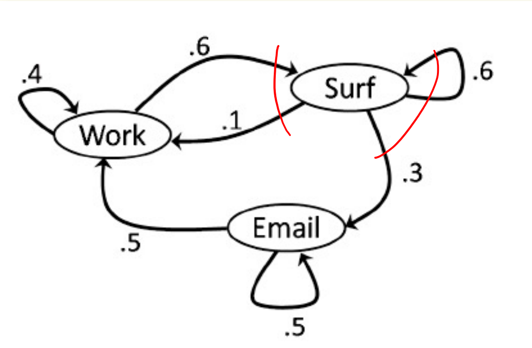
Definition. A **discrete-time stochastic process** (DTSP) is a sequence of random variables $X^{(0)}, X^{(1)}, X^{(2)}, \ldots$ where $X^{(t)}$ is the value at time t.

What happens when I start working on 312…

time $t = 0$

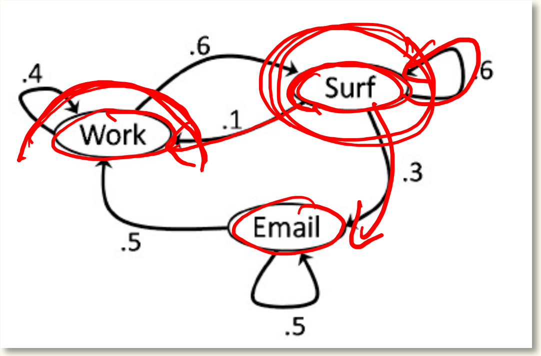






312 work habits

How do we interpret this diagram?



This kind of random process is called a **Markov Chain**

At *each* time step t

– I can be in one of 3 **states**

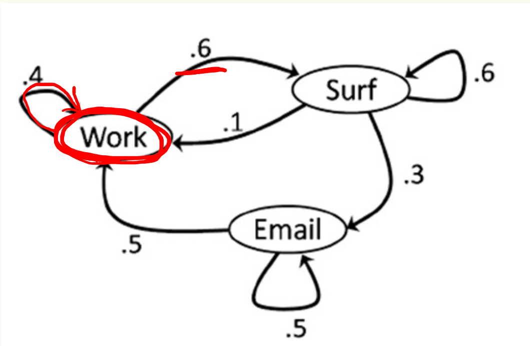
- Work, Surf, Email

– If I am in some state s at time t

- the **labels of out-edges** of s **give the probabilities** of my moving to each of the states at time $t + 1$ (as well as staying the same)
  – so **labels on out-edges sum to** 1

e.g. If I am in Email, there is a 50-50 chance I will be in each of Work or Email at the next time step, but I will never be in state Surf in the next step.

Many interesting questions about Markov Chains



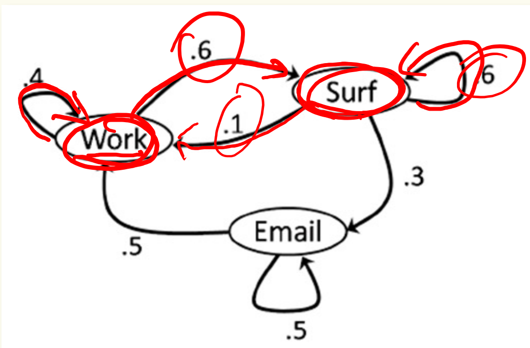
1. **What is the probability that I am in state s at time 1?**
2. **What is the probability that I am in state s at time 2?**

Define variable $X^{(t)}$ to be state I am in at time t

Given: In state Work at time $t = 0$

t	0	1	2
$P(X^{(t)} = \text{Work})$	$\mathbf{1}$	$\mathbf{0.4}$	
$P(X^{(t)} = \text{Surf})$	$\mathbf{0}$	$\mathbf{0.6}$	
$P(X^{(t)} = \text{Email})$	$\mathbf{0}$	$\mathbf{0}$	

Many interesting questions about Markov Chains



1. What is the probability that I am in state s at time 1?
2. What is the probability that I am in state s at time 2?

$\boxed{\text{Define variable } X^{(t)} \text{ to be state I am in at time } t}$

Given: In state Work at time $t = 0$

$X^{(1)}$ $X^{(2)}$ 0.1

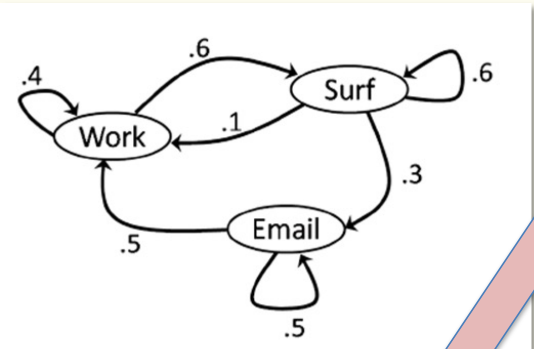
$Pr[X^1=W] \cdot Pr[X^2=W \mid X^1=W] + Pr[X^1=S] P[X^2=W \mid X^1=S]$

t	0	1	2
$q_W^{(t)} = P(X^{(t)} = \text{Work})$	$\mathbf{1}$	$\mathbf{0.4}$	$= \mathbf{0.4} \cdot 0.4 + \mathbf{0.6} \cdot 0.1 = 0.16 + 0.06 = \mathbf{0.22}$
$q_S^{(t)} = P(X^{(t)} = \text{Surf})$	$\mathbf{0}$	$\mathbf{0.6}$	$= \mathbf{0.4} \cdot 0.6 + \mathbf{0.6} \cdot 0.6 = 0.24 + 0.36 = \mathbf{0.60}$
$q_E^{(t)} = P(X^{(t)} = \text{Email})$	$\mathbf{0}$	$\mathbf{0}$	$= \mathbf{0.4} \cdot 0 + \mathbf{0.6} \cdot 0.3 = 0 + 0.18 = \mathbf{0.18}$

$Pr[X^2=S \mid X^1=W]$ $Pr[X^2=S \mid X^1=S]$

$Pr[X^1=W]$ $Pr[X^1=S]$

An organized way to understand the distribution of $X^{(t)}$

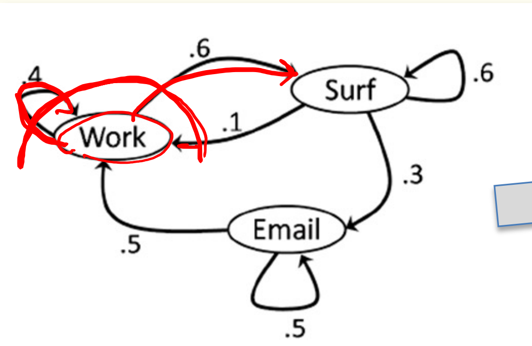


Write as a tuple $(q_W^{(t)}, q_S^{(t)}, q_E^{(t)})$ a.k.a. a row vector:

$$[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$$

t	0	1	2
$q_W^{(t)} = P(X^{(t)} = \text{Work})$	$\mathbf{1}$	$\mathbf{0.4}$	$= \mathbf{0.4} \cdot 0.4 + \mathbf{0.6} \cdot 0.1 = 0.16 + 0.06 = \mathbf{0.22}$
$q_S^{(t)} = P(X^{(t)} = \text{Surf})$	$\mathbf{0}$	$\mathbf{0.6}$	$= \mathbf{0.4} \cdot 0.6 + \mathbf{0.6} \cdot 0.6 = 0.24 + 0.36 = \mathbf{0.60}$
$q_E^{(t)} = P(X^{(t)} = \text{Email})$	$\mathbf{0}$	$\mathbf{0}$	$= \mathbf{0.4} \cdot 0 + \mathbf{0.6} \cdot 0.3 = 0 + 0.18 = \mathbf{0.18}$

An organized way to understand the distribution of $X^{(t)}$



$$M$$

$$[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \begin{array}{cc} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} \end{array} = [q_W^{(t+1)} \; q_S^{(t+1)} \; q_E^{(t+1)}]$$

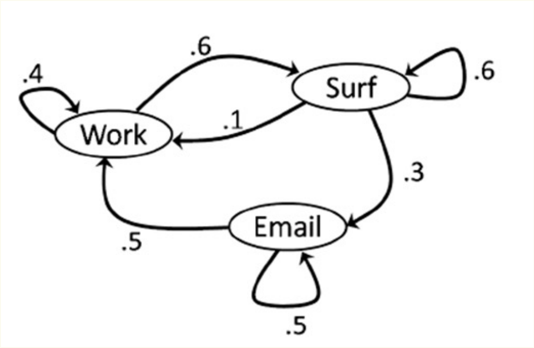
$P[\;\;|S]$

Write as a "transition probability matrix" $\boldsymbol{M}$

- one row/col per state. Row=now, Col=next
- each row sums to 1

t	0	1	2
$q_W^{(t)} = P(X^{(t)} = \text{Work})$	$\mathbf{1}$	$\mathbf{0.4}$	$= \mathbf{0.4} \cdot 0.4 + \mathbf{0.6} \cdot 0.1 = 0.16 + 0.06 = \mathbf{0.22}$
$q_S^{(t)} = P(X^{(t)} = \text{Surf})$	$\mathbf{0}$	$\mathbf{0.6}$	$= \mathbf{0.4} \cdot 0.6 + \mathbf{0.6} \cdot 0.6 = 0.24 + 0.36 = \mathbf{0.60}$
$q_E^{(t)} = P(X^{(t)} = \text{Email})$	$\mathbf{0}$	$\mathbf{0}$	$= \mathbf{0.4} \cdot 0 + \mathbf{0.6} \cdot 0.3 = 0 + 0.18 = \mathbf{0.18}$

An organized way to understand the distribution of $X^{(t)}$



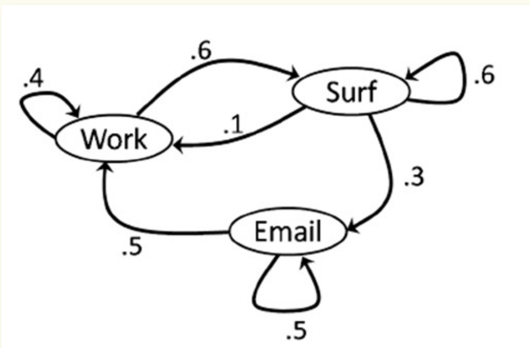
$$[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \overset{\boldsymbol{M}}{\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}} = [q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}]$$

$$q_W^{(1)} = \mathbf{0.4} \qquad q_W^{(2)} = \mathbf{0.4} \cdot 0.4 + \mathbf{0.6} \cdot 0.1 = 0.16 + 0.06 = \mathbf{0.22}$$

$$q_S^{(1)} = \mathbf{0.6} \qquad q_S^{(2)} = \mathbf{0.4} \cdot 0.6 + \mathbf{0.6} \cdot 0.6 = 0.24 + 0.36 = \mathbf{0.60}$$

$$q_E^{(1)} = \mathbf{0} \qquad q_E^{(2)} = \mathbf{0.4} \cdot 0 \quad + \mathbf{0.6} \cdot 0.3 = 0 \quad + 0.18 = \mathbf{0.18}$$

An organized way to understand the distribution of $X^{(t)}$



$$\left[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}\right] \begin{array}{c} \mathbf{M} \\ \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} \end{array} = \left[q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}\right]$$

Vector-matrix multiplication

$\Pr[X^{(t+1)} = W \mid X^{(t)} = W]$ $\quad$ $\Pr[X^{(t+1)} = W \mid X^{(t)} = S]$

$$q_W^{(t)} \cdot 0.4 + q_S^{(t)} \cdot 0.1 + q_E^{(t)} \cdot 0.5 = q_W^{(t+1)}$$

$$q_W^{(t)} \cdot 0.6 + q_S^{(t)} \cdot 0.6 + q_E^{(t)} \cdot 0 = q_S^{(t+1)}$$

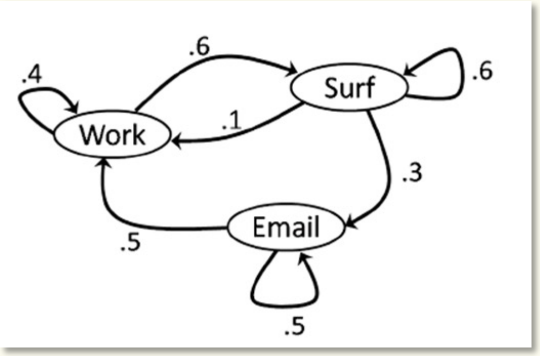
$$q_W^{(t)} \cdot 0 + q_S^{(t)} \cdot 0.3 + q_E^{(t)} \cdot 0.5 = q_E^{(t+1)}$$

$$q_W^{(1)} = \mathbf{0.4} \qquad q_W^{(2)} = \mathbf{0.4} \cdot 0.4 + \mathbf{0.6} \cdot 0.1 = 0.16 + 0.06 = \mathbf{0.22}$$

$$q_S^{(1)} = \mathbf{0.6} \qquad q_S^{(2)} = \mathbf{0.4} \cdot 0.6 + \mathbf{0.6} \cdot 0.6 = 0.24 + 0.36 = \mathbf{0.60}$$

$$q_E^{(1)} = \mathbf{0} \qquad q_E^{(2)} = \mathbf{0.4} \cdot 0 + \mathbf{0.6} \cdot 0.3 = 0 + 0.18 = \mathbf{0.18}$$

An organized way to understand the distribution of $X^{(t)}$



$$[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \overset{\boldsymbol{M}}{\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}} = [q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}]$$

Vector-matrix multiplication

$$q_W^{(t)} \cdot 0.4 + q_S^{(t)} \cdot 0.1 + q_E^{(t)} \cdot 0.5 = q_W^{(t+1)}$$

$$q_W^{(t)} \cdot 0.6 + q_S^{(t)} \cdot 0.6 + q_E^{(t)} \cdot 0 = q_S^{(t+1)}$$

$$q_W^{(t)} \cdot 0 + q_S^{(t)} \cdot 0.3 + q_E^{(t)} \cdot 0.5 = q_E^{(t+1)}$$

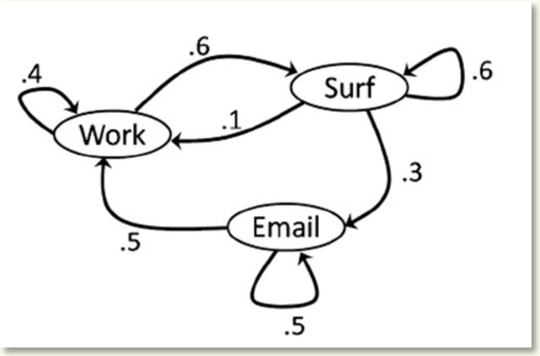
Write $\boldsymbol{q}^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$ Then for all $t \geq 0$, $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}\boldsymbol{M}$

So $\boldsymbol{q}^{(1)} = \boldsymbol{q}^{(0)}\boldsymbol{M}$

$\boldsymbol{q}^{(2)} = \boldsymbol{q}^{(1)}\boldsymbol{M} = (\boldsymbol{q}^{(0)}\boldsymbol{M})\boldsymbol{M} = \boldsymbol{q}^{(0)}\boldsymbol{M}^2$

…

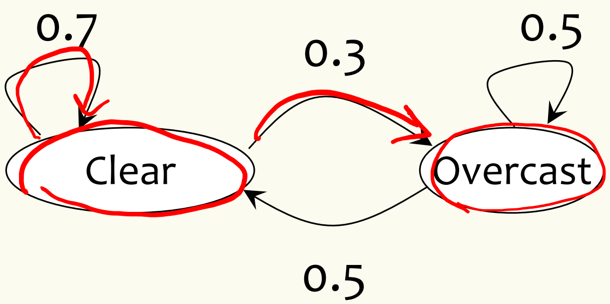
By induction … we can derive



$$\begin{array}{c} \boldsymbol{M} \\ \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} \end{array}$$

$$\boxed{\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)} \boldsymbol{M}^t \text{ for all } t \geq 0}$$

Another example:



Suppose that $\boldsymbol{q}^{(0)} = \left[q_C^{(0)}, q_O^{(0)}\right] = [0, 1]$

We have $\boldsymbol{M} = \begin{bmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{bmatrix}$

$[0, 1]\, M^2$

$[0, 1] \begin{bmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{bmatrix}$

$[0.5 \quad 0.5] \begin{bmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{bmatrix}$

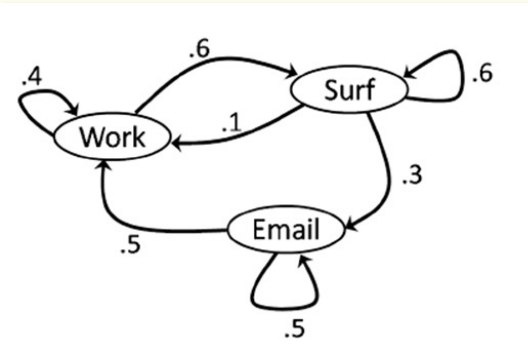
Poll: pollev.com/rachel312

What is $\boldsymbol{q}^{(2)}$?

a. $[0.3, 0.7]$
b. $[0.6, 0.4]$
c. $[0.7, 0.3]$
d. $[0.5, 0.5]$
e. $[0.4, 0.6]$

Brain Break

Many interesting questions about Markov Chains



Given: In state Work at time $t = 0$

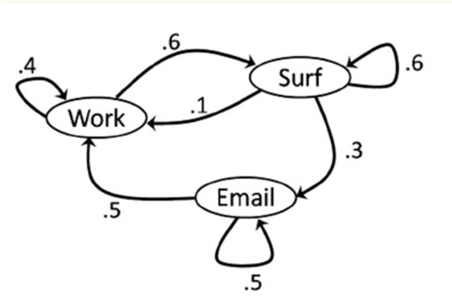
1. **What is the probability that I am in state s at time 1?**
2. **What is the probability that I am in state s at time 2?**
3. **What is the probability that I am in state s at some time t far in the future?**

$$\boxed{\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\boldsymbol{M}^t \text{ for all } t \geq 0}$$

What does $\boldsymbol{M}^t$ look like for really big t ?

$\boldsymbol{M}^t$ as t grows

$\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)} \boldsymbol{M}^t$ for all $t \geq 0$



$\boldsymbol{M}$

$$\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$\boldsymbol{M}^2$

$$\begin{array}{c c} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{pmatrix} .22 & .6 & .18 \\ .25 & .42 & .33 \\ .45 & .3 & .25 \end{pmatrix} \end{array}$$

$\boldsymbol{M}^3$

$$\begin{array}{c c} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{pmatrix} .238 & .492 & .270 \\ .307 & .402 & .291 \\ .335 & .450 & .215 \end{pmatrix} \end{array}$$

$\boldsymbol{M}^{10}$

$$\begin{array}{c c} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{pmatrix} .2940 & .4413 & .2648 \\ .2942 & .4411 & .2648 \\ .2942 & .4413 & .2648 \end{pmatrix} \end{array}$$

$\boldsymbol{M}^{30}$

$$\begin{array}{c c} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{pmatrix} .29411764705 & .44117647059 & .26470588235 \\ .29411764706 & .44117647058 & .26470588235 \\ .29411764706 & .44117647059 & .26470588235 \end{pmatrix} \end{array}$$

$\boldsymbol{M}^{60}$

$$\begin{array}{c c} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{pmatrix} .294117647058823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \end{pmatrix} \end{array}$$

What does this say about $\boldsymbol{q}^{(t)}$?

$[1, 0, 0]\, M^{60} = [\quad]$

$[0, 1, 0]\, M^{60} = [\quad]$

What does this say about $q^{(t)} = q^{(0)} M^t$?

- Note that no matter what probability distribution $q^{(0)}$ is …
$q^{(t)} =$ $q^{(0)} M^t$ is just a *weighted average* of the rows of M^t

- If every row of M^t were *exactly* the same … that would mean that $q^{(0)} M^t$ would be equal to the common row
  - So $q^{(t)}$ wouldn't depend on $q^{(0)}$

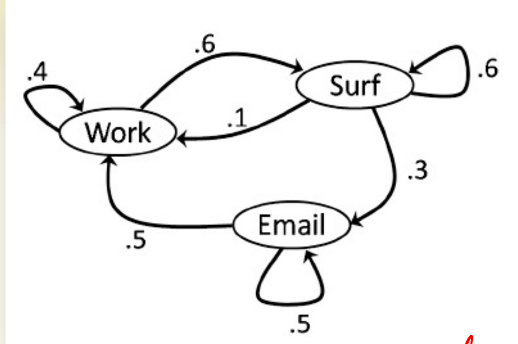
- The rows aren't exactly the same but they are very close
  - So $q^{(t)}$ barely depends on $q^{(0)}$ after very few steps

Observation

$q^{t+1} = \boxed{q^{t} M = q^{t}}$

If $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}$ then it will never change again!

$q^{t+2} = q^{t+1} M = (q^{t} M) M = q^{t} M = q^{t}$



Called a **stationary distribution** and has a special name

$\boldsymbol{\pi} = (\pi_W, \pi_S, \pi_E)$

Solution to $\boldsymbol{\pi} = \boldsymbol{\pi} M$

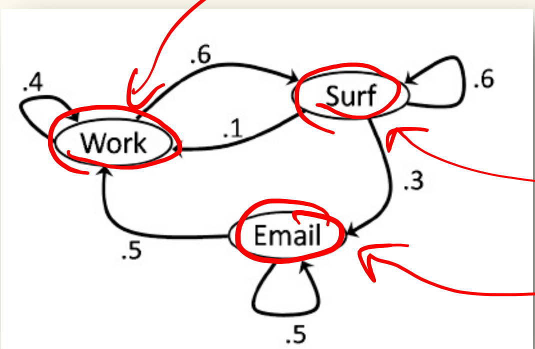
Solving for Stationary Distribution

$$M = \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$

Stationary Distribution satisfies

- $\boldsymbol{\pi} = \boldsymbol{\pi M}$, where $\boldsymbol{\pi} = (\pi_W, \pi_S, \pi_E)$
- $\pi_W + \pi_S + \pi_E = 1$

$\Rightarrow \pi_W = \frac{10}{34}, \pi_S = \frac{15}{34}, \pi_E = \frac{9}{34}$



As $t \to \infty$, $\boldsymbol{q}^{(t)} \to \boldsymbol{\pi}$ no matter what distribution $\boldsymbol{q}^{(0)}$ is !!

Markov Chains in general

- A set of n **states** $\{1, 2, 3, \ldots\ n\}$
- The state at time t is denoted by $X^{(t)}$
- A **transition matrix** $\boldsymbol{M}$, dimension $n \times n$

$$\boldsymbol{M}_{ij} = P\left(X^{(t+1)} = j \;\middle|\; X^{(t)} = i\right)$$

- $\boldsymbol{q}^{(t)} = (q_1^{(t)}, q_2^{(t)}, \ldots, q_n^{(t)})$ where $q_i^{(t)} = P(X^{(t)} = i)$
- Transition: LTP $\Rightarrow \boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}\boldsymbol{M}$ so $\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\boldsymbol{M}^t$
- A **stationary distribution** $\boldsymbol{\pi}$ is the solution to:

$$\boldsymbol{\pi} = \boldsymbol{\pi}\,\boldsymbol{M}, \text{ normalized so that } \Sigma_{i\in[n]}\pi_i = 1$$

The Fundamental Theorem of Markov Chains

Theorem. Any Markov chain that is

- irreducible* and
- aperiodic*

has a unique stationary distribution $\boldsymbol{\pi}$.

Moreover, as $t \to \infty$, for all i, j, $\boxed{\boldsymbol{M}_{ij}^{t}} = \pi_j$

$+ \uparrow \uparrow$

$$q^{(0)} \cdot M^{t} = \begin{bmatrix} \pi \\ \pi \\ \pi \\ \pi \end{bmatrix} = \pi$$

$$\pi = \pi M$$

These concepts are way beyond us but they turn out to cover a very large class of Markov chains of practical importance.