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**

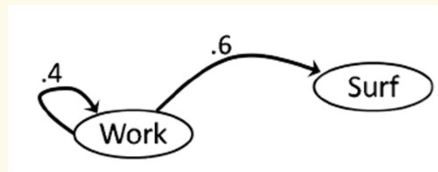
More generally: randomness can enter over many steps and depend on previous outcomes

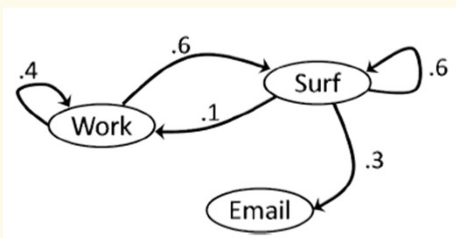
Random Process 1 → **Outcome Distribution** D_1 → Random Process 2 → **Outcome Distribution** D_2 → Random Process 3 → **Outcome Distribution** D_3 …

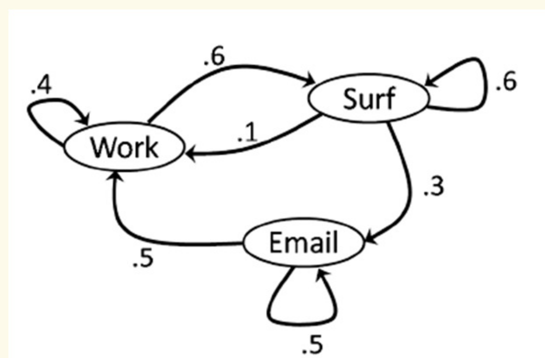
Definition. A **discrete-time stochastic process** (DTSP) is a sequence of random variables $X^{(0)}, X^{(1)}, X^{(2)}, \ldots$ where $X^{(t)}$ is the value at time t.

What happens when I start working on 312…

time $t = 0$

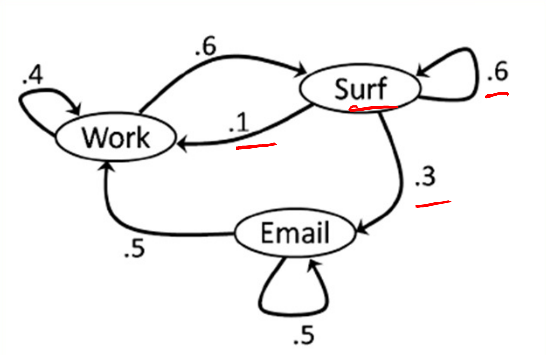






312 work habits

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

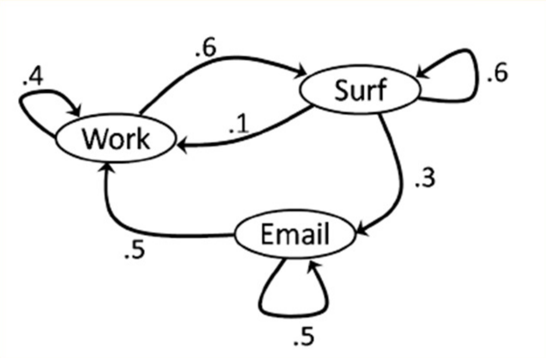
– If I am in some state s at time t

- 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.

This diagram looks vaguely familiar if you took CSE 311 …

Markov chains are a special kind of *probabilistic (finite) automaton*

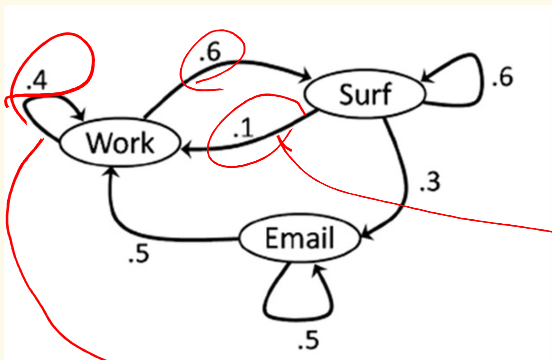


The diagrams look a bit like those of Deterministic Finite Automata (DFAs) you saw in 311 except that…

- There are no input symbols on the edges
  - Think of there being only one kind of input symbol "another tick of the clock" so no need to mark it on the edge
- They have multiple out-edges like an NFA, except that they come with probabilities

But just like DFAs, the only thing they remember about the past is the state they are currently in.

Many interesting questions about Markov Chains



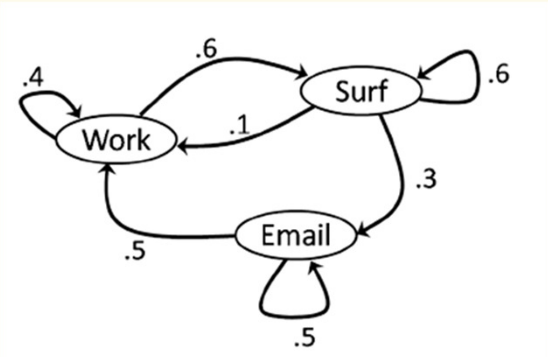
1. What is the probability that I am in state s at time 1?
2. What is the probability that I am in state s at time 2?

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
$P(X^{(t)} = \text{Work})$	**1**	**0.4**	$= 0.4 \times 0.4 + 0.6 \times 0.1 = 0.16 + .06 = 0.22$
$P(X^{(t)} = \text{Surf})$	**0**	**0.6**	
$P(X^{(t)} = \text{Email})$	**0**	**0**	

Many interesting questions about Markov Chains



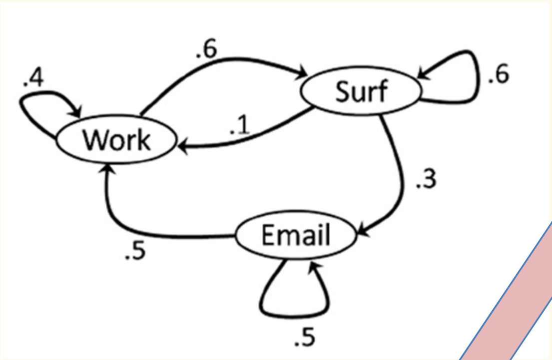
1. **What is the probability that I am in state s at time 1?**
2. **What is the probability that I am in state s at time 2?**

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

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

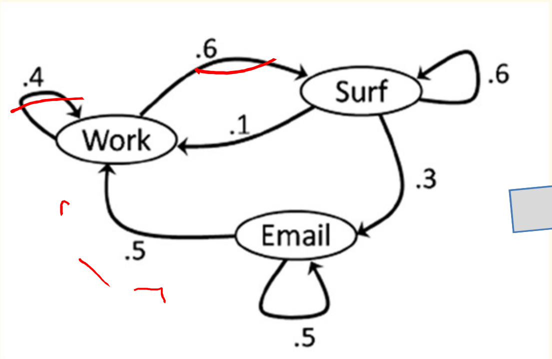


Write as a tuple $(q_W^{(t)}, q_S^{(t)}, q_E^{(t)})$ a.k.a. a row vector:

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

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 $X^{(t)}$



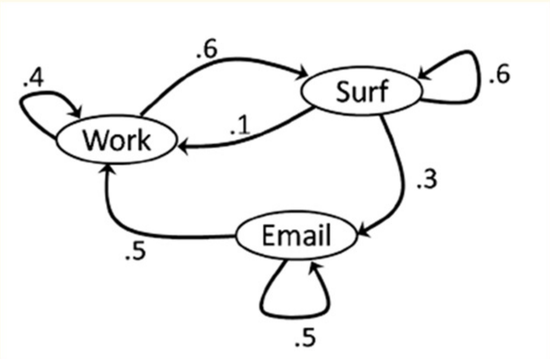
$$[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \quad \boldsymbol{M} = \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

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 $X^{(t)}$



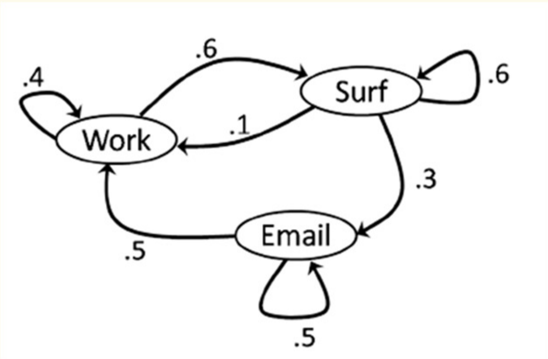
$$\left[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}\right] \overset{\boldsymbol{M}}{\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}} = \left[q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}\right]$$

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

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

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

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



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

Vector-matrix multiplication

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

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

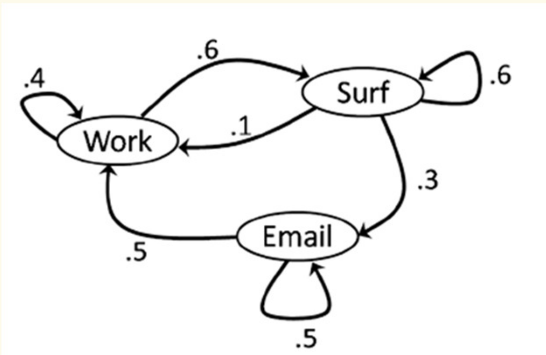
$$q_W^{(t)} \cdot 0 + q_S^{(t)} \cdot 0.3 + q_E^{(t)} \cdot 0.5 = q_E^{(t+1)}$$

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

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

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

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

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

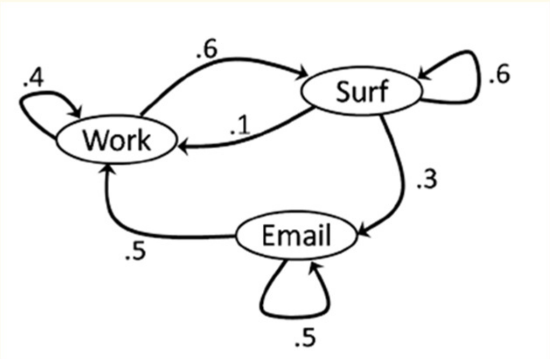
…

associativity

matrix multiplication

$M \cdot M$

By induction … we can derive

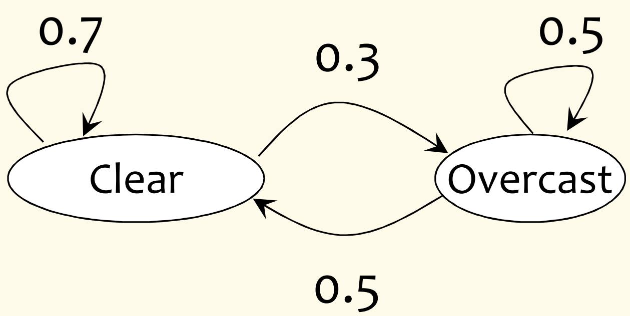


$$\boldsymbol{M}$$

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

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

Another example:



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

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

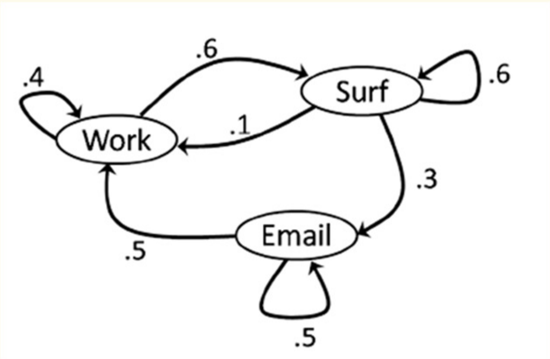
Poll: pollev.com/paulbeame028

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 $t = 0$

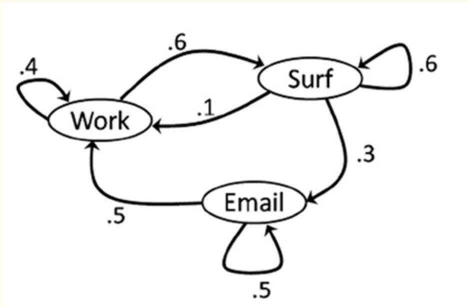
1. **What is the probability that I am in state s at time 1?**
2. **What is the probability that I am in state s at time 2?**
3. **What is the probability that I am in state s at some time t far in the future?**

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

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

M^t as t grows

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



M

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

M^2

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

M^3

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

M^{10}

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

M^{30}

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

M^{60}

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

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

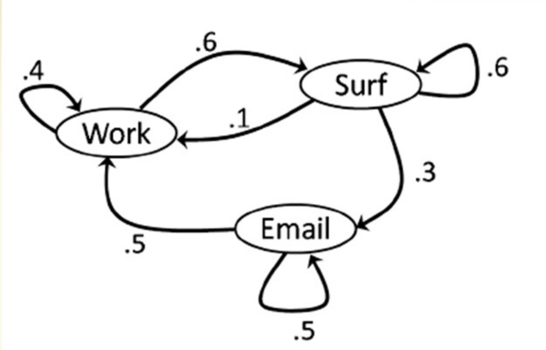
What does this say about $\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\boldsymbol{M}^t$?

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

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

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

Observation



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

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

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

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

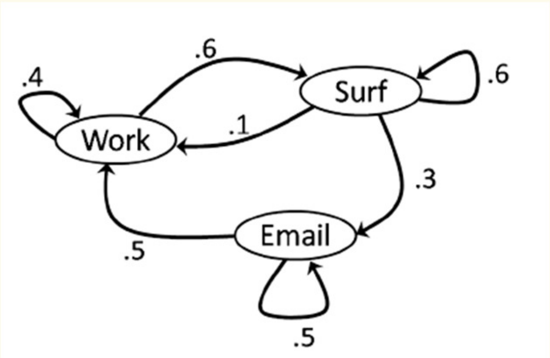
Solving for Stationary Distribution

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

Stationary Distribution satisfies

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

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



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

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

- $\boldsymbol{q}^{(t)} = (q_1^{(t)}, q_2^{(t)}, \dots, q_n^{(t)})$ 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, $\boldsymbol{M}_{ij}^{t} = \pi_j$

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