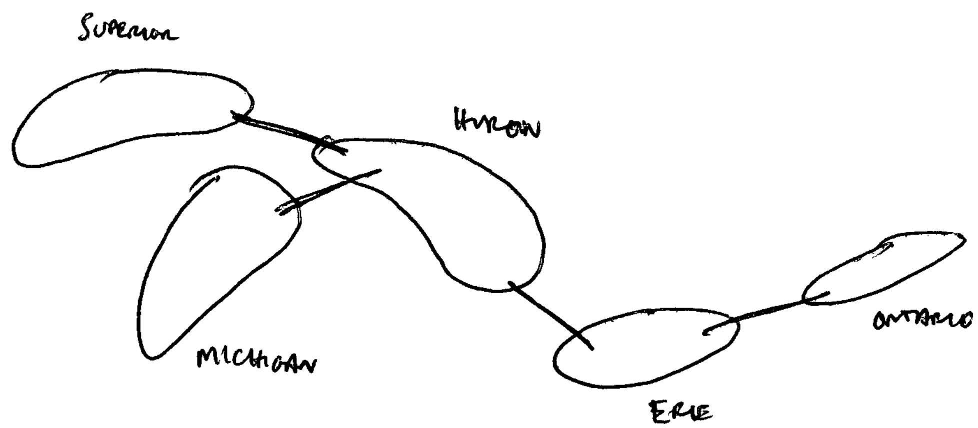
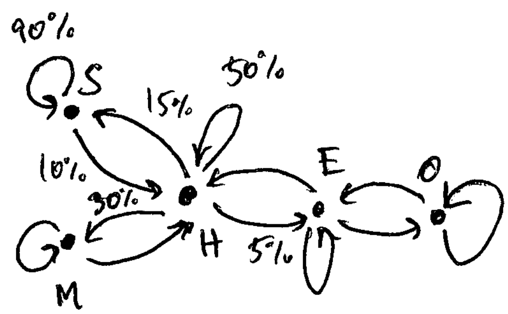
LECTURE 20
03/05/12

MARKOV CHAINS AND STOCHASTIC MATRICES (5.7):

EX.





STATE SPACE = $\{S, M, H, E, O\}$.

MODEL: EVERY YEAR, A FIXED PROPORTION OF FISH MIGRATE FROM EACH GREAT LAKE TO ANOTHER LAKE.

WE ASSUME

(i) THESE PROPORTIONS STAY THE SAME YEAR BY YEAR

(ii) WE KNOW THE INITIAL POPULATION OF FISH IN EACH LAKE.

LET A_{ij} = PROPORTION THAT GO FROM LAKE j TO LAKE i IN ONE YEAR.

THEN,

$$A = \begin{array}{c} \begin{array}{ccccc} S & M & H & E & O \end{array} \\ \underbrace{\begin{pmatrix} .9 & \cdot & .15 & \cdot & \cdot \\ 0 & \cdot & .3 & \cdot & \cdot \\ .1 & \cdot & .5 & \cdot & \cdot \\ 0 & 1 & .5 & \cdot & \cdot \\ 0 & \cdot & 0 & \cdot & \cdot \end{pmatrix}}_{\text{FROM}} \end{array} \begin{array}{c} S \\ M \\ H \\ E \\ O \end{array} \Bigg\} \text{TO}$$

NOTE: COLUMNS OF A SUM TO 1!

DEF. A IS A (LEFT) STOCHASTIC TRANSITION MATRIX IF $\underline{r} A = \underline{r}$, $\underline{r} = (1\ 1 \cdots 1)$ AND ALL ENTRIES OF A ARE NONNEGATIVE $(A_{ij} \geq 0)$.

$\underline{r}$ ← ROW VECTOR!

- THEN, $(A^k)_{ij} = \sum_{l_1} \cdots \sum_{l_{k-1}} A_{il_1} A_{l_1 l_2} \cdots A_{l_{k-1} j}$

IS THE PROPORTION THAT MOVE FROM j TO i IN EXACTLY k STEPS, GIVEN STARTED AT j.

($A_{il_1} \cdots A_{l_{k-1}j}$ CORRESPONDS TO PROPORTION THAT MOVE ALONG PATH $j \to l_{k-1} \to l_{k-2} \to \cdots \to l_1 \to i$.)

DEF. $\underline{v}$ IS A PROBABILITY VECTOR IF $\underline{r}\,\underline{v} = 1$ AND ALL ENTRIES OF $\underline{v}$ ARE NONNEGATIVE $(\underline{v}_i \geq 0)$.

⟶ WE DENOTE SPACE OF ALL PROBABILITY VECTORS (DISTRIBUTIONS) BY $\mathcal{P}$.

- IF INITIALLY THE PROPORTION OF FISH IN LAKE i IS $\underline{v}_i$, THEN k YEARS LATER IT IS $(A^k \underline{v})_i$.

PROBABILISTIC INTERPRETATION:

- A_{ij} IS THE TRANSITION PROBABILITY $p_{ij} = P(X_1 = i \mid X_0 = j)$ THAT AN INDIVIDUAL FISH IS AT STATE i AT TIME 1 GIVEN IT IS AT STATE j AT TIME 0 (X_k IS THE POSITION AT TIME k).

- "PROBABILITY" = PROPORTIONS OBSERVED OVER MANY INDEPENDENT TRIALS (THINK OF COIN TOSSES).
  THIS A CONSEQUENCE OF THE SO-CALLED LAW OF LARGE NUMBERS (FAIR COIN HAS 50% PROBABILITY OF H OR T ON ANY ONE FLIP SINCE OVER MANY REPEATED FLIPS WE OBSERVE 50% SHOW UP H OR T IN THE LONG-RUN).

- $\underline{x}(k) = A^k \underline{x}(0)$ IS A MARKOV CHAIN WITH INITIAL DISTRIBUTION $\underline{x}(0)$ (A PROBABILITY VECTOR).

  $\underline{x}(k)$ = PROBABILITY ANY ONE FISH IS IN A PARTICULAR STATE AT TIME $k \geq 0$
  = PROPORTION OF ENSEMBLE OF ALL FISH THAT IS IN PARTICULAR STATE AT TIME $k \geq 0$.

- MARKOV BECAUSE TRANSITION PROBABILITIES ONLY DEPEND ON CURRENT STATE, NOT ON THE PAST.
  FOR EX., IF FISH DO NOT RETURN TO THEIR PREVIOUS STATE IMMEDIATELY, THE MODEL IS NOT MARKOV.

Lecture 21
03/07/12

PLEASE CONSULT NOTES ON FOLLOWING PAGES, ALONG WITH DISCUSSION ON A PARTICULARLY IMPORTANT EXAMPLE OF A MARKOV CHAIN (RANDOM WALK ON FINITE SET OF STATES), FOR THE REMAINDER OF OUR DISCUSSION ON MARKOV CHAINS.

:: Review ::
-State space (set of nodes), probability vectors v \in P (distribution on state space), stochastic matrix A (entries are transition probabilities for an individual, become proportions that transition when considering large ensembles by LLN).
-Markov property: transition probabilities only depend on current state, not on history of path taken to get there. For ex., if fish model modified such that fish will not return to previous lake right away, not Markov! Models with finite memory can be made Markov by enlarging the state space.

:: Properties of A (Perron-Frobenius theorem) ::
(0) A^k is transition matrix. In particular, if x(0) \in P then x(k) = (A^k)x(0) \in P.
Pf.: A^k has all positive entries and r(A^k) = r(A^{k-1}) = ... = r.

-Progressively draw spectrum with each step.
(1) A always has eigenvalue 1 (possibly with multiplicity greater than 1). Corresponding eigenvectors can be normalized to be in P.
Pf.: rA = r => A^T has eigenvalue 1 => A has eigenvalue 1. *Eigenvector in P not shown.*

(2) All eigenvalues of A must lie in closed unit disc of C (i.e., A has no eigenvalues of magnitude greater than 1, or A has spectral radius 1). Corresponding eigenvectors must have entires sum to 1.
Pf.: First, A cannot have eigenvalue corresponding to unstable mode since otherwise x(k) --> \infty, which contradicts (0). Second, (r - rA) = 0 => 0 = (r - rA)\xi = (1 - \lambda)r\xi => r\xi = 0.

(3) If A has all positive entries, then 1 is only eigenvalue on unit circle in C, and has algebraic multiplicity 1.
Pf.: See Q3 for proof. *Algebraic multiplicity 1 not shown.*

:: Stationary distributions ::
-Def. Stationary distribution is \pi \in P such that A\pi = \pi.
-We are interested in stationary distributions because they are statistical equilibria of the system (for example, temperature in a room settles down to a fixed profile that depends on distance from floor--warmest air on top, coolest on bottom due to gravity). If system is in a statistically stationary state note that the random state of any one individual is *not* fixed in time, but the distribution of states is.

Questions:

*Q0: When does a stationary distribution \pi exist?
A: Always by (1). In fact, if x(k) \to v \in P, must be to a stationary distribution (v = \pi).

*Q1: Is \pi unique?
  A: Not necessarily. As we have seen, eigenvalue 1 can have algebraic multiplicity greater than 1. Counterexample: A = I means every probability vector is a stationary distribution. Problem is that that two sets of states never communicate with each other (can't get from one set of states to other). For example, A = [block 1; ... ; block N] also allows for nonuniqueness. To overcome this, we impose *irreducibility* of A---for each fixed i,j, (A^k)_ij > 0 for some k (i.e., can eventually get from every state to every other state).

*Q2: If \pi unique, does x(k) = (A^k)x(0) converges to \pi for every x(0)?
  A: Not necessarily. Counterexample: A = [0 1; 1 0]. Then every x(0) \in P besides x(0) = (1/2, 1/2)^T does not converge. Problem is periodicity. More generally, A = shift of identity also allows for periodicity, or A transition matrix of periodic random walk on even number of states. To overcome this, need *aperiodicity* of A---for each fixed i, there is a K such that (A^k)_ii > 0 for all k \geq K (i.e., returns to state i do not form a rigid pattern).

-Picture: Venn diagram. Irreducibility (uniqueness of \pi) \cap aperiodicity (convergence) = regularity (convergence to unique stationary distribution \pi). To deal with Q1, Q2, we impose irreducibility and aperiodicity. Equivalently, this is the condition of *regularity* of transition matrix.
-Def. A is regular if for some K \geq 1, A^K has all entries strictly positive (that is, all states communicate in at most K steps). Then A^k has positive entries for all k \geq K. Can show that A is regular iff it is irreducible and aperiodic (HW problem).

-Theorem. For regular A, we have a unique \pi to which every initial state converges. In addition, A^k converges to (\pi ... \pi).
Pf.: Assuming A regular with K = 1 WLOG, uniqueness and convergence by (1)-(3). Consider x(0) = e_i for each i to get A^k --> (\pi ... \pi).

*Q3: If A regular, how to find \pi? How fast does algorithm converge?
  A: Power method. Start with any initial condition x(0), and evolve. Converges at rate given by second eigenvalue \lambda_2. For a regular matrix A with K = 1, this can be estimated by |\lambda_2| \leq (1 - n*min(A)) (in general, one has |\lambda_2|^K \leq (1 - n*min(A))).
Pf.: Eigenvector v corresponding to \lambda is also eigenvector of A - min(A)*B = (1 - n*min(A))\tilde{A} for B = [r ; ... ; r] and \tilde{A} a stochastic matrix. But since \tilde{A} has all eigenvalues with magnitude less than 1, must have |\lambda| \leq (1 - n*min(A)) < 1. In particular, this implies that convergence must be at least as fast at (1 - n*min(A))^k.

*Q4: What do these distributional properties about the ensemble imply about any particular random path?
  A: Ergodicity--long-run time average of any chosen path equals \pi, which is the long-run ensemble average at a fixed time. In other words, for a regular Markov chain each path is representative of the entire ensemble. True even for periodic transition matrices (still need irreducibility in order get a unique \pi).
-Theorem. \lim_{T --> \infty} (1/T)*\sum_1^T x(k) = \pi.

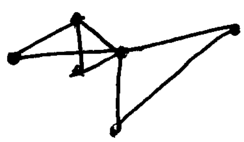
Pf.: [?]

*Q5: What if we drop irreducibility? Can we still get unique \pi?
  A: Sometimes. For example, A = [1 1; 0 0] is transition matrix for an absorbing Markov chain. But if we had more than one absorbing state, this wouldn't be true (why?). In fact, nonuniqueness for stochastic matrices of form A = [I B; 0 C] (absorbing Markov chains with absorbing states in I), where I has dimension greater than 1.

LECTURE 22
03/09/12

APPLICATIONS TO NETWORK SCIENCE:

<u>NETWORK</u>: SET S OF n <u>NODES</u>, CONNECTED BY <u>EDGES</u>.



<u>UNDIRECTED</u>

(E.G., FACEBOOK)

<u>DIRECTED</u>.

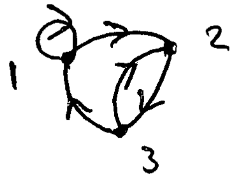
(E.G., TWITTER, WWW).

WE WILL FOCUS ON DIRECTED NETWORKS.

<u>DEF.</u> ADJACENCY MATRIX A OF NETWORK GIVEN BY

$$A_{ij} = \begin{cases} 1 & \text{IF THERE IS A DIRECTED EDGE FROM } j \text{ TO } i \\ 0 & \text{ELSE} \end{cases} \qquad (j \to i)$$

<u>EX.</u>



$$\Longrightarrow \quad A = \begin{array}{c} \text{(FROM)} \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \end{array} \begin{array}{l} \\ \\ 1 \\ 2 \text{ (TO)} \\ 3 \end{array}$$

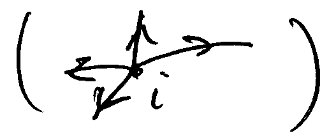
<u>NOTE</u>: (i) $A_{ij} \geq 0$ FOR ALL i, j.

(ii) IF NETWORK <u>UNDIRECTED</u>, CAN CONSIDER EACH EDGE AS A PAIR OF DIRECTED EDGES, SO A IS SYMMETRIC.

ALL PROPERTIES OF NETWORK CAN BE DIRECTLY OBTAINED FROM A. FOR EXAMPLE:

(i) IN-DEGREE OF NODE i: $d_i^{in} \doteq \sum_{j=1}^{n} A_{ij}$ (ROW SUM)

(ii) OUT-DEGREE OF NODE i: $d_i^{out} = \sum_{j=1}^{n} A_{ji}$ (COLUMN SUM)



(iii) # OF PATHS OF LENGTH 2 FROM j TO i:

$$N^{(2)}_{ij} \doteq \sum_{k=1}^{n} \underbrace{A_{ik} A_{kj}}_{= \begin{cases} 1 & \text{IF } j \to k \to i \\ 0 & \text{ELSE} \end{cases}} = (A^2)_{ij} .$$

TO FROM

OF PATHS OF LENGTH $r \geq 1$ FROM j TO i:

$$N^{(r)}_{ij} = (A^r)_{ij} .$$

(iv) A CYCLE OF LENGTH $r \geq 1$ IS ANY PATH OF LENGTH r THAT BEGINS AND ENDS AT THE SAME NODE.

OF CYCLES OF LENGTH $r \geq 1$ IN NETWORK:

$$C^{(r)} = \sum_{i=1}^{n} N^{(r)}_{ii} = \sum_{i=1}^{n} (A^r)_{ii} = \mathrm{Tr}(A^r) .$$

SINCE A HAS JORDAN FORM $A = P\tilde{D}P^{-1}$,

$$Tr(A^r) = Tr(P\tilde{D}^r P^{-1}) = Tr(\tilde{D}^r)$$
$$= \lambda_1^r + \dots + \lambda_n^r$$

WHERE $\lambda_1, \dots, \lambda_n$ ARE THE EIGENVALUES OF A.

NOTE: WHEN COUNTING CYCLES, WE DISTINGUISH THE EXACT ORDER OF EDGES TRAVERSED. THAT IS,

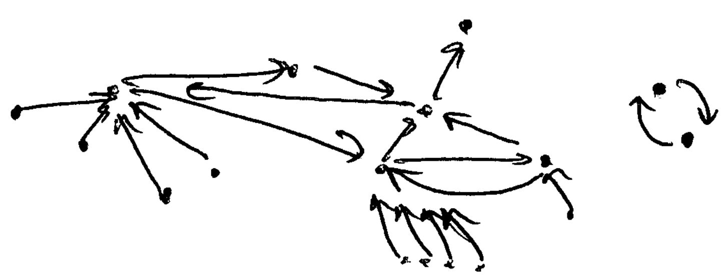
HAS 3 CYCLES OF LENGTH 3

$(1 \to 2 \to 3 \to 1,\ 2 \to 3 \to 1 \to 2,$
AND $3 \to 1 \to 2 \to 3)$.

PAGERANK / EIGENVECTOR CENTRALITY:

Q: GIVEN A DIRECTED NETWORK, HOW IMPORTANT ("CENTRAL") IS NODE i?

EX. (WWW)



- NODES ARE WEBPAGES, DIRECTED EDGES ARE LINKS.

IDEA: LET x_i DENOTE THE IMPORTANCE (CENTRALITY) OF WEBPAGE i. DENOTE $\underline{x} = (x_1, \dots, x_n)^T$.

- CENTRALITY x_i SHOULD DEPEND ON
  - (i) # OF PAGES THAT LINK TO i
  - (ii) HOW CENTRAL THESE LINKING PAGES ARE THEMSELVES (IMPORTANCE BEGETS IMPORTANCE).
  - (iii) HOW MANY OTHER PAGES THESE LINKING PAGES LINK TO

$$\Rightarrow \quad x_i = \sum_{j=1}^{n} \frac{\overset{\text{(i)}}{A_{ij}}}{\underset{\text{(iii)}}{d_j^{out}}} \overset{\text{(ii)}}{x_j}$$

LET $T_{ij} = \dfrac{A_{ij}}{d_j^{out}}$ $\Rightarrow$ $T_{ij} \geq 0$ FOR ALL i, j AND $(\underline{1}^T T)_i = \sum_{j=1}^{n} T_{ji} = 1$ SINCE d_j^{out} TERM NORMALIZES A_{ij}.

$\Rightarrow T$ TRANSITION MATRIX.

SO, CENTRALITY $\underline{x}$ SATISFIES

$$\underline{x} = T \underline{x},$$

I.E., STATIONARY DISTRIBUTION OF MARKOV CHAIN DETERMINED BY T!

NOTE THAT SUCH A STATIONARY DIST. $\underline{x} \in P$ ALWAYS EXISTS, BUT MAY NOT BE UNIQUE SINCE T IS NOT NECESSARILY REGULAR (THAT IS, IT MAY BE REDUCIBLE OR PERIODIC)!

TO REMEDY THIS, LET'S GIVE EVERY NODE A LITTLE BIT OF CENTRALITY FOR "FREE":

$$x_i = \underset{\substack{\uparrow \\ 0 \le \alpha \le 1 \\ \text{DAMPING FACTOR}}}{\alpha} \Big(\underbrace{\sum_{j=1}^{n} \frac{A_{ij}}{d_j^{out}} x_j}_{\text{FROM BEFORE}} \Big) + (1-\alpha) \underbrace{\frac{1}{n}}_{\text{FOR "FREE"}}$$

$$\Rightarrow \underline{x} = \alpha T \underline{x} + (1-\alpha) \frac{1}{n} \underline{r}^T \qquad \underline{r} = (1 \; 1 \cdots 1). \text{ ROW VECTOR.}$$

NOTE:

$$\underline{r}\,\underline{x} = \alpha \underbrace{\underline{r}\,T}_{=\underline{r}} \underline{x} + (1-\alpha)\frac{1}{n} \underbrace{\underline{r}\,\underline{r}^T}_{=n}$$

$$= \alpha \underline{r}\,\underline{x} + (1-\alpha)$$

$$\Rightarrow (1-\alpha)\underline{r}\,\underline{x} = (1-\alpha) \Rightarrow \underline{r}\,\underline{x} = 1 \quad \text{IF } \alpha \neq 1.$$

SO, FOR $\alpha \neq 1$, $\underline{x} \in P$ AND WE CAN WRITE

$$\underline{x} = \alpha T \underline{x} + (1-\alpha)\frac{1}{n}\, \underline{r}^T(\underline{r}\,\underline{x})$$

$$= \left[\alpha T + (1-\alpha)\frac{1}{n}\underbrace{\underline{r}^T\underline{r}}_{= \begin{pmatrix}1 & \cdots & 1\\ \vdots & \ddots & \vdots\\ 1 & \cdots & 1\end{pmatrix}}\right]\underline{x}$$

CALL $T_{PR} = \alpha T + (1-\alpha)\frac{1}{n}\underline{r}^T\underline{r}$.

THIS IS A <u>REGULAR</u> TRANSITION MATRIX SINCE THE SMALLEST ENTRY IS AT LEAST AS LARGE AS $(1-\alpha)\frac{1}{n}$!

- T_{PR} CAN BE THOUGHT OF AS AN INTERPOLATION (I.E., A MIXTURE) OF THE TRANSITION MATRIX T WITH THE TRANSITION MATRIX $\frac{1}{n}\underline{r}^T\underline{r}$, WITH $\alpha < 1$ DETERMINING THE LEVEL OF MIXING.

<u>DEF.</u> <u>PAGE RANK</u> OF i IS $\underline{\pi}_i$, WHERE $\underline{\pi} \in P$ IS THE <u>UNIQUE</u> SOLN. OF $\underline{x} = T_{PR}\,\underline{x}$.

THIS IS THE ORIGINAL ALGORITHM EMPLOYED BY GOOGLE TO RANK WEBPAGES, WITH α TAKEN TO BE 0.85.

REMARK:

1) STARTING WITH ANY $\underline{x}(0) \in P$ AS AN INITIAL GUESS,

$$\underline{x}(k) = T_{PR}^{k}\, \underline{x}(0) \xrightarrow{k\to\infty} \underline{\pi}$$

2) T_{PR} IS TRANSITION MATRIX OF RANDOM WALK ON WEBPAGES (W/ PROB. α OF FOLLOWING ONE OF THE LINKS UNIFORMLY AT RANDOM VS. PROB. $1-\alpha$ OF JUMPING TO A RANDOMLY CHOSEN WEBPAGE, AT EACH STEP). THAT IS, $\underline{x}(k)$ IS THE PROBABILITY THAT A "RANDOM SURFER" IS AT A GIVEN WEBPAGE AFTER k STEPS, AND $\underline{\pi}$ IS ITS LIMIT AS $k \to \infty$.

3) COULD REPLACE TERM $\frac{1}{n}\underline{r}^T$ BY SOME $\beta \in P$ TO WEIGHT IMPORTANCE BASED ON <u>CONTEXT</u> (I.E., RANDOM SURFER JUMPS TO PAGE WITH DISTRIBUTION GIVEN BY β INSTEAD OF UNIFORMLY AT RANDOM.)

WE COMPUTE $\underline{\pi}$ USING THE <u>POWER METHOD</u> (SINCE IT IS TYPICALLY IMPRACTICAL TO FIND THE EIGENVECTORS OF A LARGE MATRIX LIKE T_{PR} FOR $n \gg 1$!):

POWER METHOD:

① CHOOSE INITIAL GUESS FOR PAGERANK, SAY $\underline{x}(0) \in P$.

② ITERATE ⟵ $\underline{x}(k) = T_{PR}^{k}\, \underline{x}(0)$.

FOR k LARGE ENOUGH, $\underline{x}(k) \approx \underline{\pi}$.

- RATE OF CONVERGENCE:

$$|\lambda_2| \leq \Big(1 - n \underbrace{\min_{ij} (T_{PR})_{ij}}_{\geq (1-\alpha)\frac{1}{n}}\Big)$$

$$\leq \Big(1 - n \cdot \frac{(1-\alpha)}{n}\Big)$$

$$\leq \boxed{\alpha} \longleftarrow \text{DAMPING FACTOR.}$$

SO, $\|\underline{x}(k) - \underline{\pi}\| \leq \text{CONST.} \times |\lambda_2|^k \leq \text{CONST.} \times \alpha^k$.

DEPENDS ON $\underline{x}(0)$.

- (1998) PAGE-BRIN: 322 MILLION LINKS

⇝ CONVERGENCE IN ≈ 52 ITERATIONS.