Multi-agent learning

Emergence of Conventions

Gerard Vreeswijk, Intelligent Systems Group, Computer Science Department, Faculty of Sciences, Utrecht University, The Netherlands.

Tuesday 16th January, 2018

Emergence of conventions: motivation

Emergence of conventions: motivation

- Many forms of social and economic behavior evolve from the bottom up: language, manners, protocols, dress code, money and credit.

Emergence of conventions: motivation

- Many forms of social and economic behavior evolve from the bottom up: language, manners, protocols, dress code, money and credit.
- For the most part no one dictated their form; they simply emerged.

Emergence of conventions: motivation

- Many forms of social and economic behavior evolve from the bottom up: language, manners, protocols, dress code, money and credit.
- For the most part no one dictated their form; they simply emerged.
- We will discuss H. Peyton Young's framework for analyzing such systems based on massive perturbed Markov chains (1993).

Emergence of conventions: motivation

- Many forms of social and economic behavior evolve from the bottom up: language, manners, protocols, dress code, money and credit.
- For the most part no one dictated their form; they simply emerged.
- We will discuss H. Peyton Young's framework for analyzing such systems based on massive perturbed Markov chains (1993).
  - Agents act rationally most of the time

Emergence of conventions: motivation

- Many forms of social and economic behavior evolve from the bottom up: language, manners, protocols, dress code, money and credit.
- For the most part no one dictated their form; they simply emerged.
- We will discuss H. Peyton Young's framework for analyzing such systems based on massive perturbed Markov chains (1993).
  - Agents act rationally most of the time, but sometimes they do not for whatever reason (error of perception, idiosyncrasy, invention, experimentation).

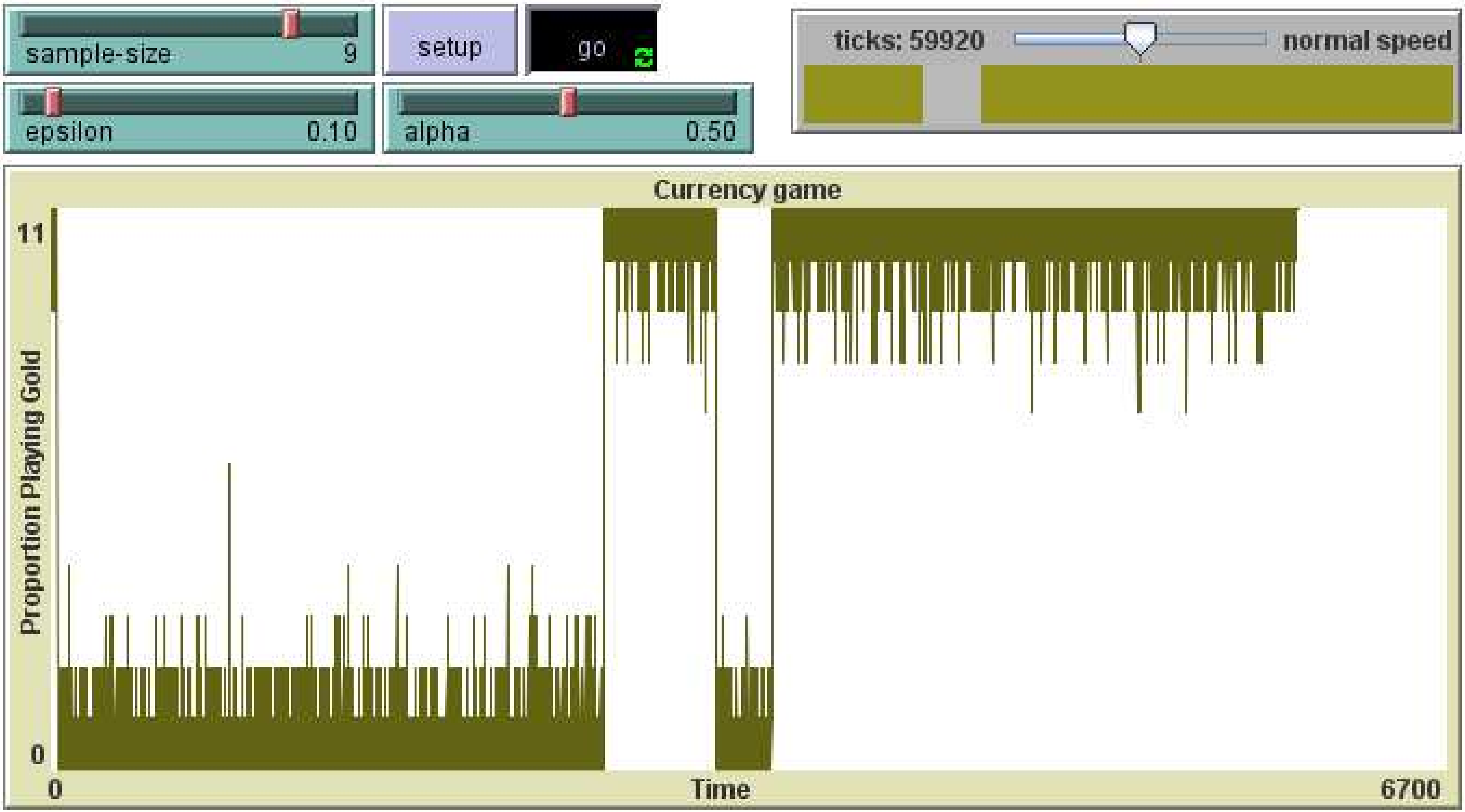
Emergence of conventions: motivation

- Many forms of social and economic behavior evolve from the bottom up: language, manners, protocols, dress code, money and credit.
- For the most part no one dictated their form; they simply emerged.
- We will discuss H. Peyton Young's framework for analyzing such systems based on massive perturbed Markov chains (1993).
  - Agents act rationally most of the time, but sometimes they do not for whatever reason (error of perception, idiosyncrasy, invention, experimentation).
  - The presence of perturbations implies that the dynamic never settles but remains in flux.

Emergence of conventions: motivation

- Many forms of social and economic behavior evolve from the bottom up: language, manners, protocols, dress code, money and credit.
- For the most part no one dictated their form; they simply emerged.
- We will discuss H. Peyton Young's framework for analyzing such systems based on massive perturbed Markov chains (1993).
  - Agents act rationally most of the time, but sometimes they do not for whatever reason (error of perception, idiosyncrasy, invention, experimentation).
  - The presence of perturbations implies that the dynamic never settles but remains in flux.
  - If minimal random behaviour is allowed, "sub-optimal" stable states ascend into "global optimal" stable states.

Example: the emergence of common currency



Plan for today

Plan for today

1. **Markov processes.** (Ergodic process, communicating states/class, transient state/class, recurrent state/class, periodic state/class, absorbing state, irreducible process, stationary distribution.)

   Compute stationary distributions:

Plan for today

1. **Markov processes.** (Ergodic process, communicating states/class, transient state/class, recurrent state/class, periodic state/class, absorbing state, irreducible process, stationary distribution.)

   Compute stationary distributions:

   - Solve n linear equations.

Plan for today

1. **Markov processes.** (Ergodic process, communicating states/class, transient state/class, recurrent state/class, periodic state/class, absorbing state, irreducible process, stationary distribution.)

   Compute stationary distributions:

   - Solve n linear equations.
   - Compare n so-called z-trees (Freidlin and Wentzell, 1984).

Plan for today

1. **Markov processes.** (Ergodic process, communicating states/class, transient state/class, recurrent state/class, periodic state/class, absorbing state, irreducible process, stationary distribution.)

   Compute stationary distributions:

   - Solve n linear equations.
   - Compare n so-called z-trees (Freidlin and Wentzell, 1984).

2. **Perturbed Markov processes.** (Regular perturbed Markov process, punctuated equilibrium, stochastically stable state.)

   Compute stochastically stable states:

Plan for today

1. **Markov processes.** (Ergodic process, communicating states/class, transient state/class, recurrent state/class, periodic state/class, absorbing state, irreducible process, stationary distribution.)

   Compute stationary distributions:

   - Solve n linear equations.
   - Compare n so-called z-trees (Freidlin and Wentzell, 1984).

2. **Perturbed Markov processes.** (Regular perturbed Markov process, punctuated equilibrium, stochastically stable state.)

   Compute stochastically stable states:

   - Compare k so-called z-trees, where k is the number of Recurrence classes in the non-perturbed Markov process (Peyton Young, 1993).

Plan for today

3. **Applications.**

Plan for today

3. **Applications.**

- Emergence of a currency standard.

Plan for today

3. **Applications.**

- Emergence of a currency standard.
- Competing technologies: operating system A vs. operating system B.

Plan for today

3. **Applications.**

- Emergence of a currency standard.
- Competing technologies: operating system A vs. operating system B.
- Competing technologies: cell phone company A vs. cell phone company B. (If time allows.)

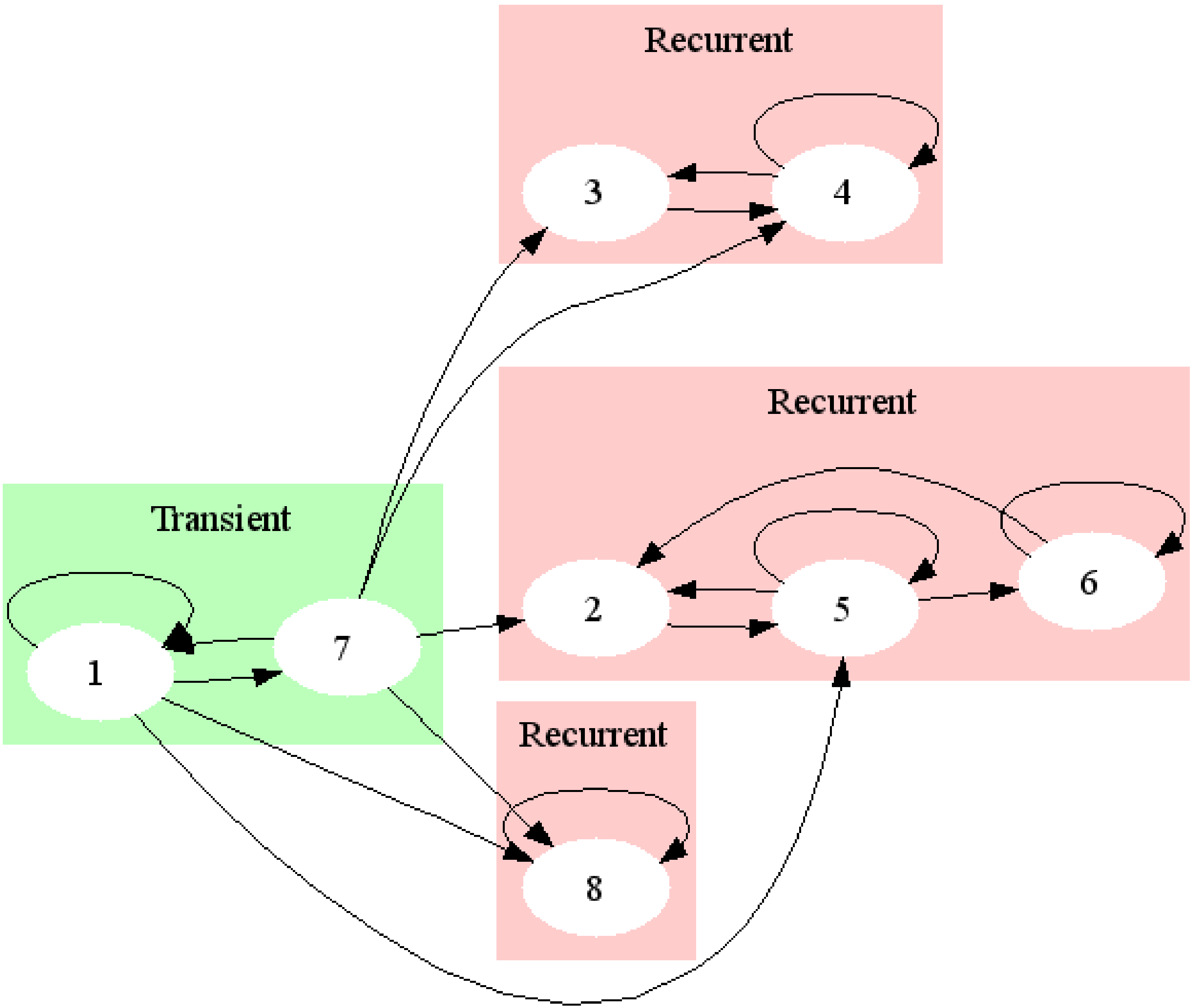
Plan for today

3. **Applications.**

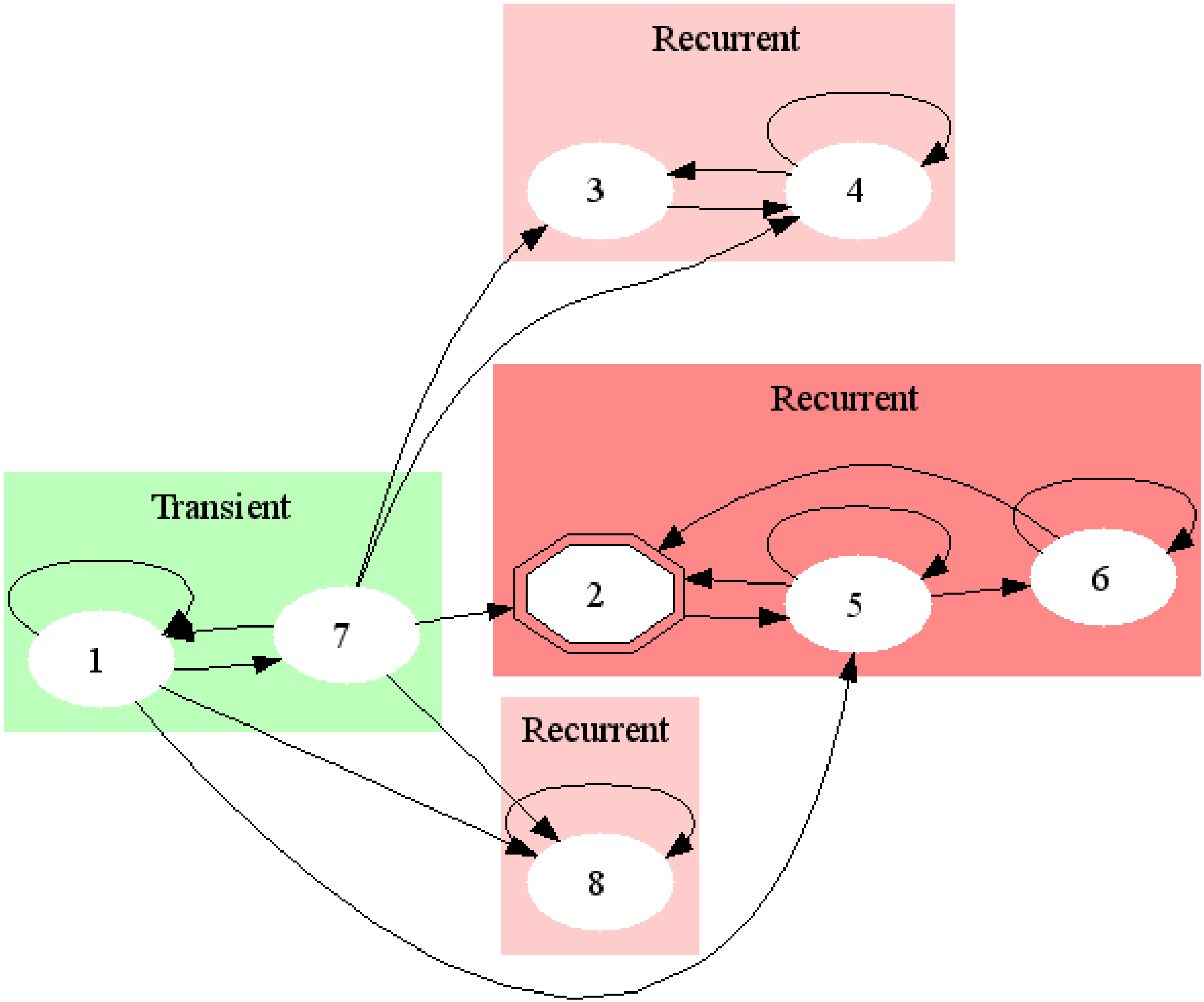
- Emergence of a currency standard.
- Competing technologies: operating system A vs. operating system B.
- Competing technologies: cell phone company A vs. cell phone company B. (If time allows.)
- Schelling's model of segregation (1969).

Part 1: Markov processes

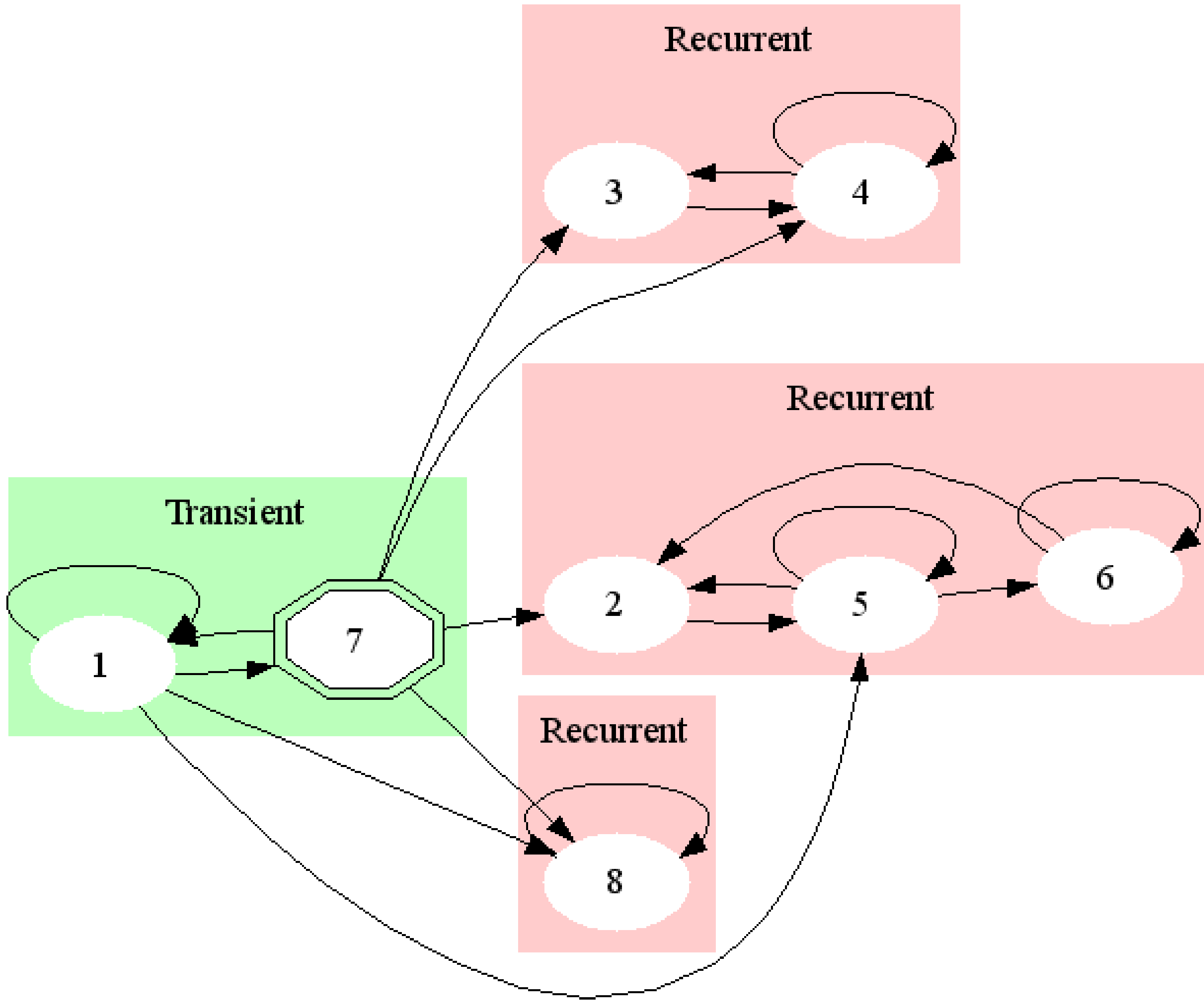
Communication classes



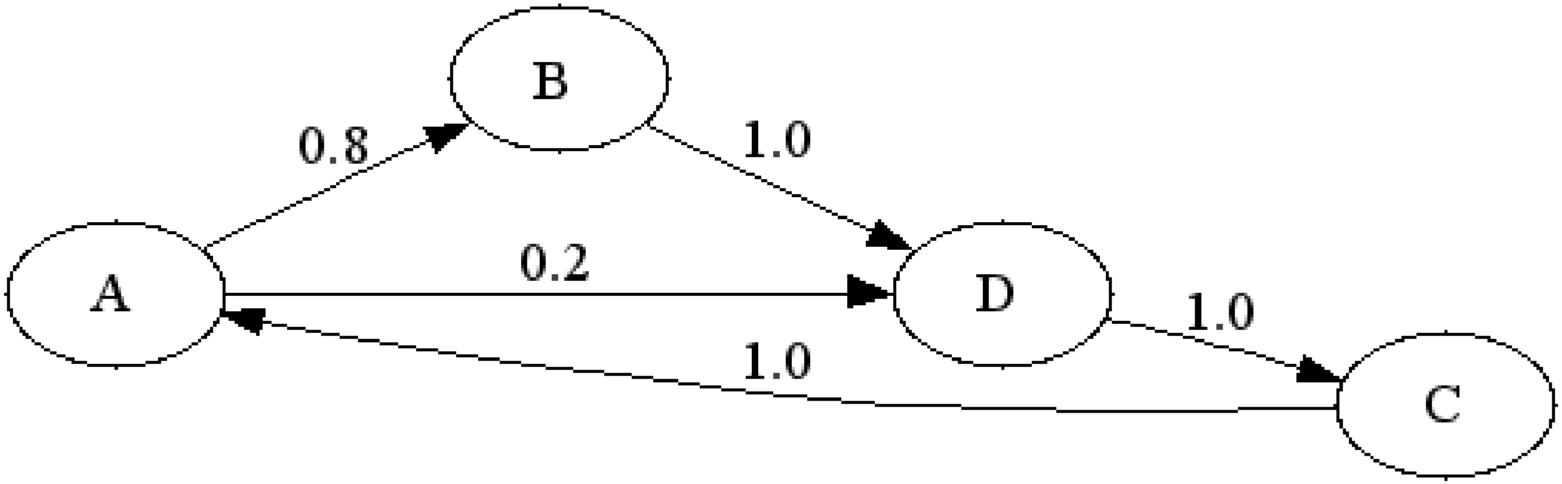
Start state matters



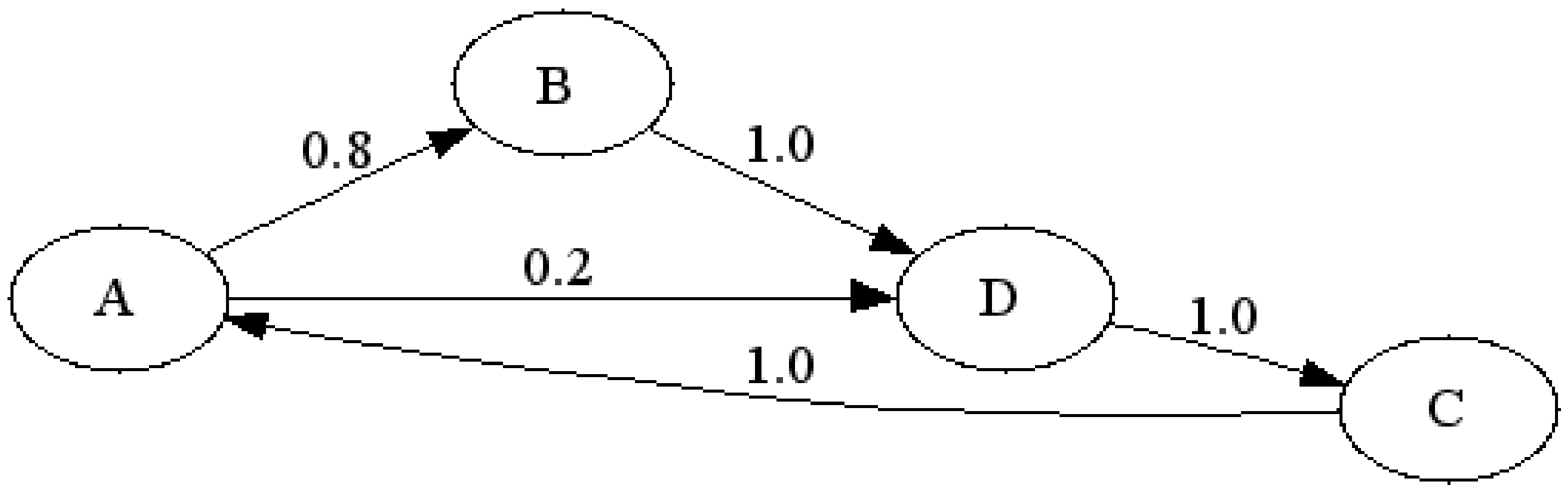
Start state matters... but here it does not



Computing the stationary distribution with linear algebr



Computing the stationary distribution with linear algebr

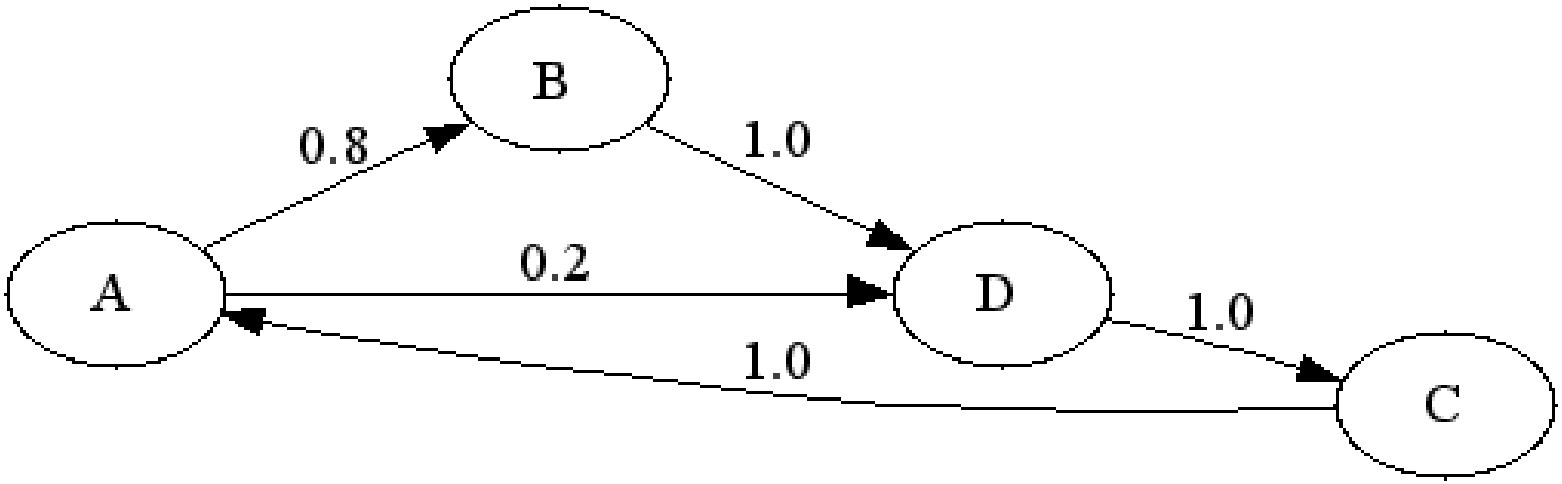


$$P(A) = P(A|A')P(A') + P(A|B')P(B') + P(A|C')P(C') + P(A|D')P(D')$$

Suppose that visiting probabilities are stationary ($A = A'$, $B = B'$, $\dots$):

$$\begin{aligned} &= P(A|A)P(A) + P(A|B)P(B) + P(A|C)P(C) + P(A|D)P(D) \\ &= 0 \cdot P(A) + 0 \cdot P(B) + 1 \cdot P(C) + 0 \cdot P(D) = P(C). \end{aligned}$$

Computing the stationary distribution with linear algebr



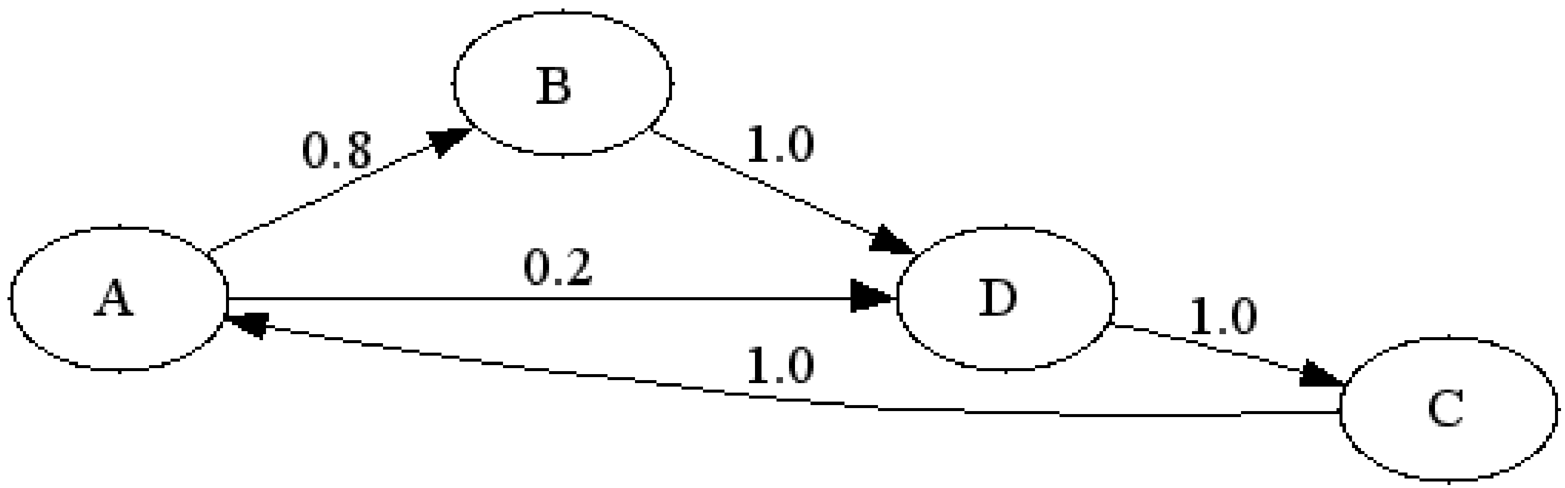
$$P(A) = P(A|A')P(A') + P(A|B')P(B') + P(A|C')P(C') + P(A|D')P(D')$$

Suppose that visiting probabilities are stationary ($A = A'$, $B = B'$, $\dots$):

$$\begin{aligned} &= P(A|A)P(A) + P(A|B)P(B) + P(A|C)P(C) + P(A|D)P(D) \\ &= 0 \cdot P(A) + 0 \cdot P(B) + 1 \cdot P(C) + 0 \cdot P(D) = P(C). \end{aligned}$$

Let us write this as $A = C$. Similarly, $B = 0.8A$, $C = D$, and $D = 0.2A + B$.

Computing the stationary distribution with linear algebr



$$P(A) = P(A|A')P(A') + P(A|B')P(B') + P(A|C')P(C') + P(A|D')P(D')$$

Suppose that visiting probabilities are stationary ($A = A'$, $B = B'$, $\ldots$):

$$\begin{aligned} &= P(A|A)P(A) + P(A|B)P(B) + P(A|C)P(C) + P(A|D)P(D) \\ &= 0 \cdot P(A) + 0 \cdot P(B) + 1 \cdot P(C) + 0 \cdot P(D) = P(C). \end{aligned}$$

Let us write this as $A = C$. Similarly, $B = 0.8A$, $C = D$, and $D = 0.2A + B$. Four (five) equations with four unknowns.

Theory of discrete Markov processes

Definitions:

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.
- Process is path-dependent: empirical distribution depends on start state. Ergodic otherwise.

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.
- Process is path-dependent: empirical distribution depends on start state. Ergodic otherwise.
- Node is recurrent: will be revisited a.s. Transient otherw.

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.
- Process is path-dependent: empirical distribution depends on start state. Ergodic otherwise.
- Node is recurrent: will be revisited a.s. Transient otherw.
- Process is irreducible: all states can reach each other.

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.
- Process is path-dependent: empirical distribution depends on start state. Ergodic otherwise.
- Node is recurrent: will be revisited a.s. Transient otherw.
- Process is irreducible: all states can reach each other.

Facts:

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.
- Process is path-dependent: empirical distribution depends on start state. Ergodic otherwise.
- Node is recurrent: will be revisited a.s. Transient otherw.
- Process is irreducible: all states can reach each other.

Facts:

- Recurrence and transience are class properties.

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.
- Process is path-dependent: empirical distribution depends on start state. Ergodic otherwise.
- Node is recurrent: will be revisited a.s. Transient otherw.
- Process is irreducible: all states can reach each other.

Facts:

- Recurrence and transience are class properties.
- If finite number of states:

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.
- Process is path-dependent: empirical distribution depends on start state. Ergodic otherwise.
- Node is recurrent: will be revisited a.s. Transient otherw.
- Process is irreducible: all states can reach each other.

Facts:

- Recurrence and transience are class properties.
- If finite number of states:
  - At least one recurrence class.

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.
- Process is path-dependent: empirical distribution depends on start state. Ergodic otherwise.
- Node is recurrent: will be revisited a.s. Transient otherw.
- Process is irreducible: all states can reach each other.

Facts:

- Recurrence and transience are class properties.
- If finite number of states:
  - At least one recurrence class.
  - If precisely one recurrence class then ergodic, and conversely.

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.
- Process is path-dependent: empirical distribution depends on start state. Ergodic otherwise.
- Node is recurrent: will be revisited a.s. Transient otherw.
- Process is irreducible: all states can reach each other.

Facts:

- Recurrence and transience are class properties.
- If finite number of states:
  - At least one recurrence class.
  - If precisely one recurrence class then ergodic, and conversely.
- A stationary distribution always exists.

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.
- Process is path-dependent: empirical distribution depends on start state. Ergodic otherwise.
- Node is recurrent: will be revisited a.s. Transient otherw.
- Process is irreducible: all states can reach each other.

Facts:

- Recurrence and transience are class properties.
- If finite number of states:
  - At least one recurrence class.
  - If precisely one recurrence class then ergodic, and conversely.
- A stationary distribution always exists. Unique iff ergodic. In that case, stationary distr. $\equiv$ empirical distr.

Theory of discrete Markov processes

Definitions:

- Stationary distribution: fixed point of transition probabilities.
- Empirical distribution: long run normalised frequency of visits.
- Limit distribution: long run probability to visit a node.
- Process is path-dependent: empirical distribution depends on start state. Ergodic otherwise.
- Node is recurrent: will be revisited a.s. Transient otherw.
- Process is irreducible: all states can reach each other.

Facts:

- Recurrence and transience are class properties.
- If finite number of states:
  - At least one recurrence class.
  - If precisely one recurrence class then ergodic, and conversely.
- A stationary distribution always exists. Unique iff ergodic. In that case, stationary distr. $\equiv$ empirical distr.
- If ergodic **and** a-periodic, then stationary distr. $\equiv$ limit distr.

Finding stationary distributions with many states

Finding stationary distributions with many states

- Solve n equations in n unknowns. What if S is large?

$$\begin{pmatrix} 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\ 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\ 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\ 0.0 & 0.1 & 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\ 0.5 & 0.2 & 0.0 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.2 \\ 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\ 0.0 & 0.1 & 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\ 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\ 0.3 & 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.0 & 0.0 & 0.3 & 0.0 \\ 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \end{pmatrix}$$

Finding stationary distributions with many states

- Solve n equations in n unknowns. What if S is large?

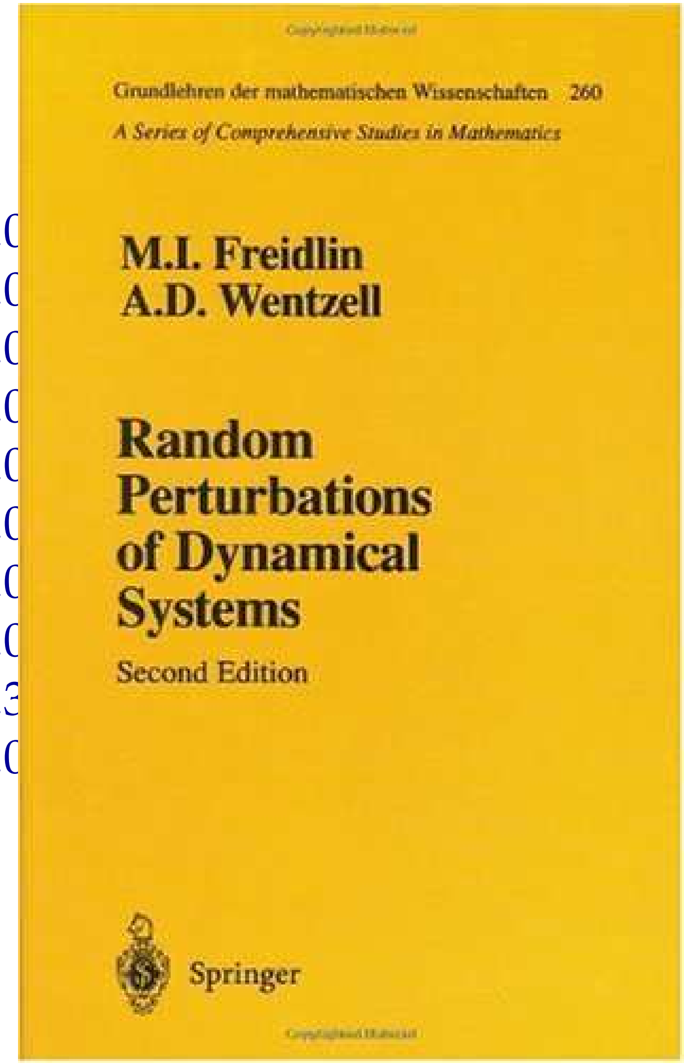
$$\begin{pmatrix}
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\
0.0 & 0.1 & 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\
0.5 & 0.2 & 0.0 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.2 \\
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\
0.0 & 0.1 & 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2 \\
0.3 & 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.0 & 0.0 & 0.3 & 0.0 \\
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 & 0.2
\end{pmatrix}$$

- Freidlin & Wentzell (1984): only look at so-called `state trees`.

Finding stationary distributions with many states

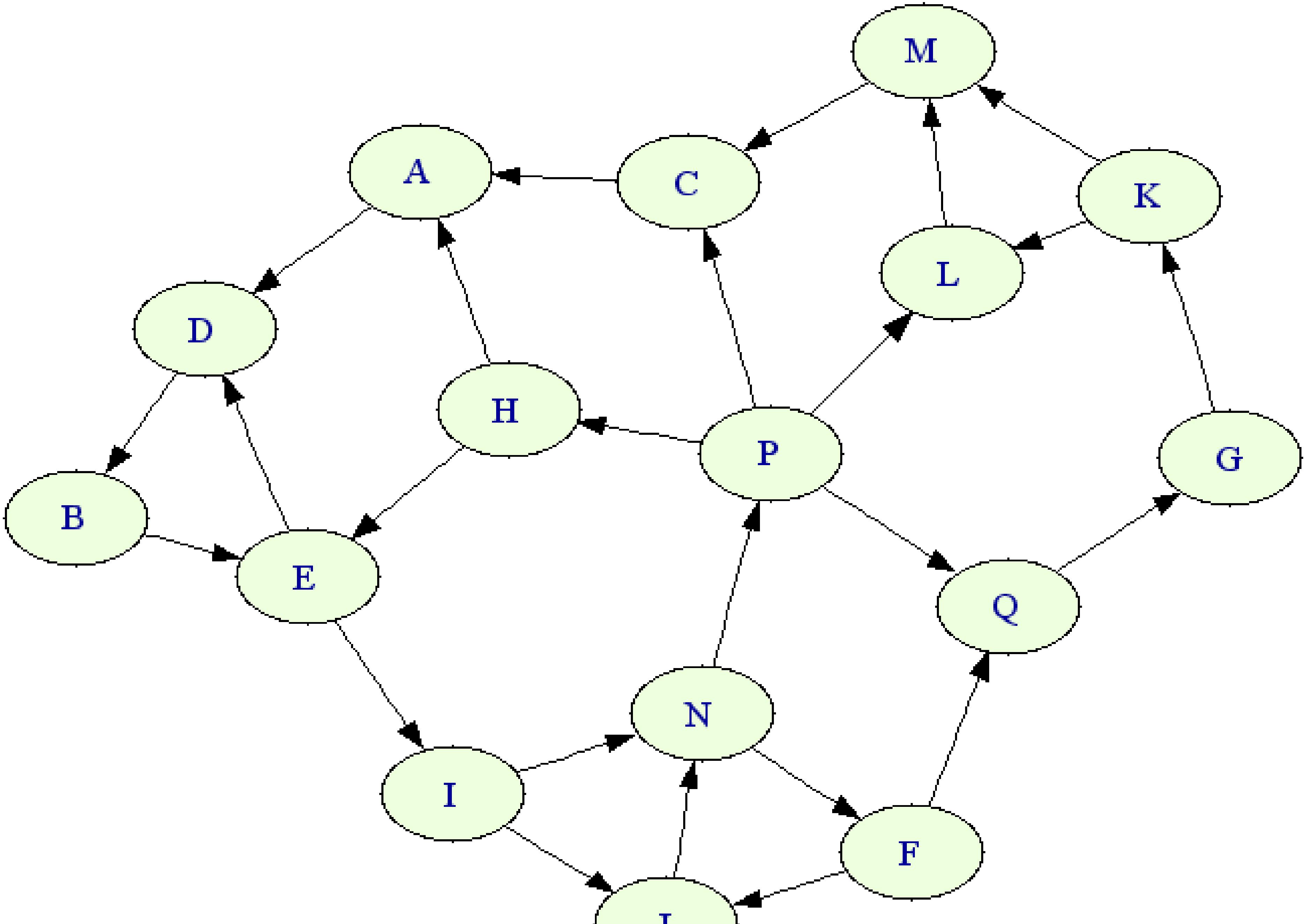
- Solve n equations in n unknowns. What if S is large?

$$\left(\begin{matrix}
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 \\
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 \\
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 \\
0.0 & 0.1 & 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 \\
0.5 & 0.2 & 0.0 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 \\
0.0 & 0.1 & 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 \\
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0 \\
0.3 & 0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.0 & 0.0 & 0.3 \\
0.1 & 0.2 & 0.0 & 0.1 & 0.0 & 0.1 & 0.0 & 0.3 & 0.0
\end{matrix}\right.$$

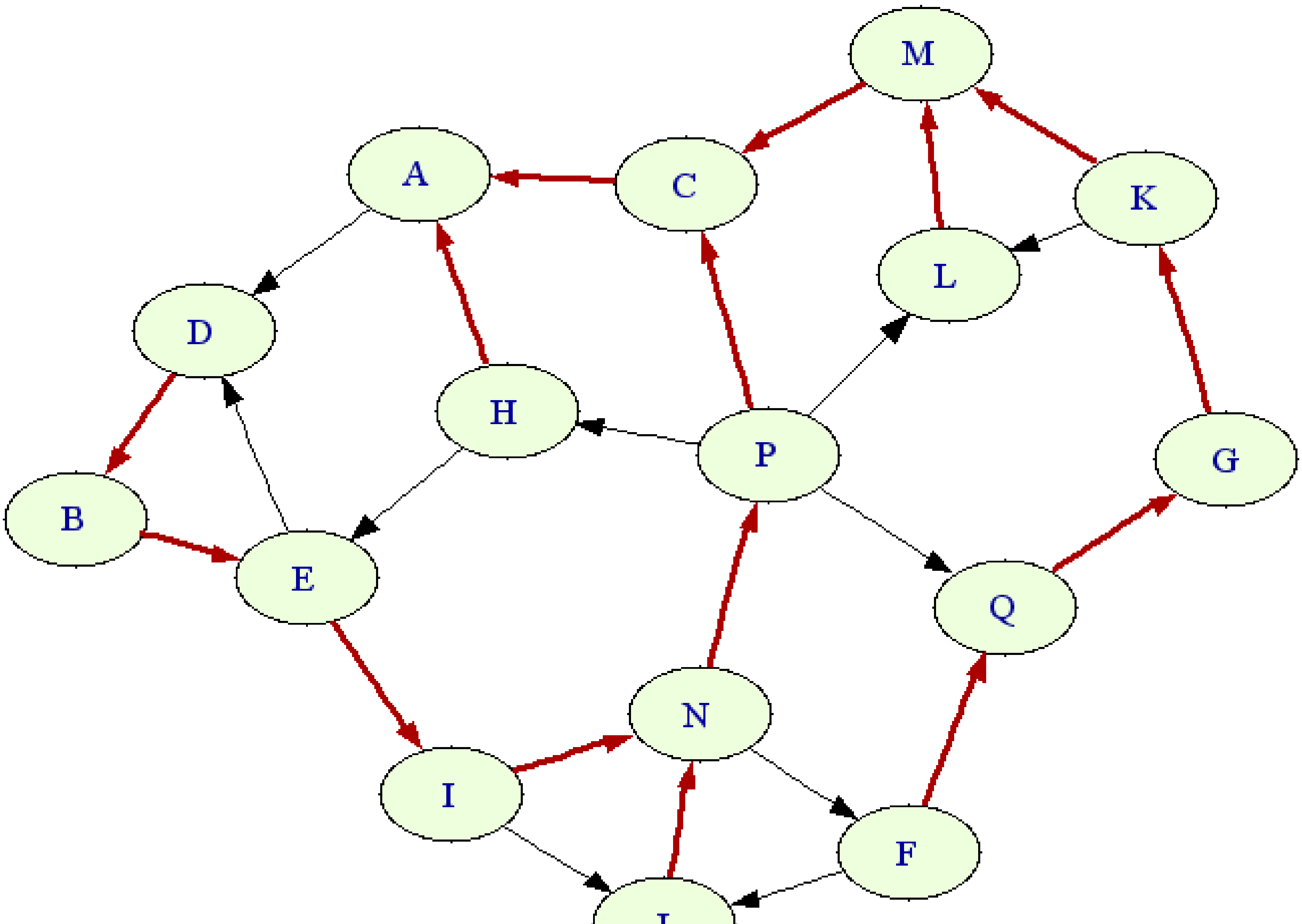


- Freidlin & Wentzell (1984): only look at so-called `state trees`.

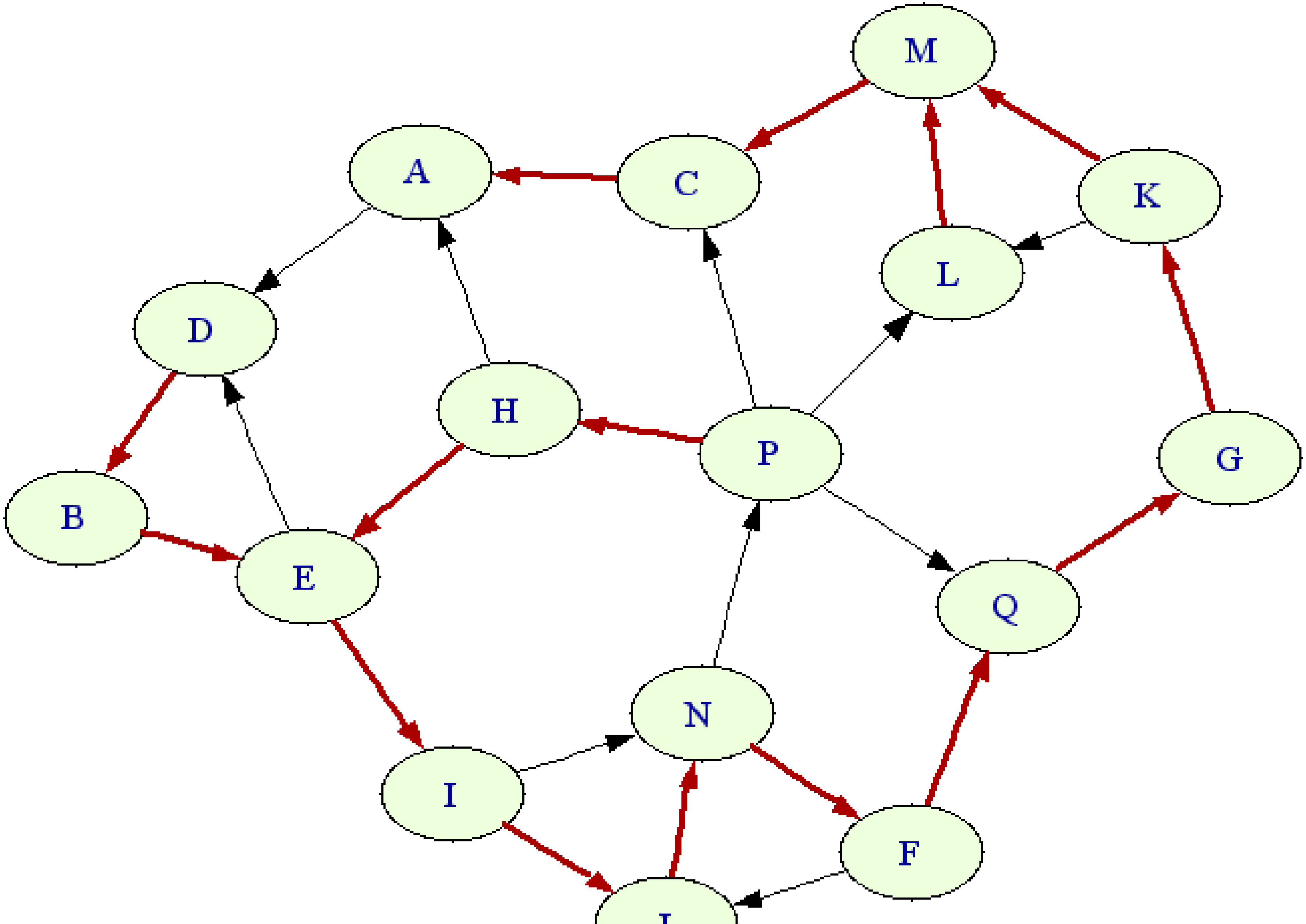
An irreducible (and finite) Markov process



One possible A-tree (i.e., tree for state A)



Another possible A-tree



An easier way to compute the stationary distribution

An easier way to compute the stationary distribution

Definition.

An easier way to compute the stationary distribution

Definition.

- An s-tree, T_s, is a complete collection of disjoint paths from states $\neq s$ to s.

An easier way to compute the stationary distribution

Definition.

- An s-tree, T_s, is a complete collection of disjoint paths from states $\neq s$ to s.
- The likelihood of an s-tree T_s, written $\ell(T_s)$, $=_{Def}$ the product of its edge probabilities.

An easier way to compute the stationary distribution

Definition.

- An s-tree, T_s, is a complete collection of disjoint paths from states $\neq s$ to s.
- The likelihood of an s-tree T_s, written $\ell(T_s)$, $=_{Def}$ the product of its edge probabilities.
- The likelihood of a state s, written $\ell(s)$, $=_{Def}$ sum of the likelihood of all s-trees.

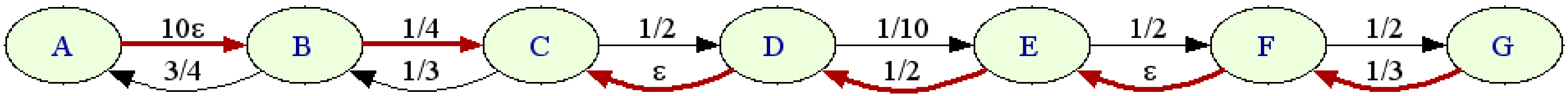
An easier way to compute the stationary distribution

Definition.

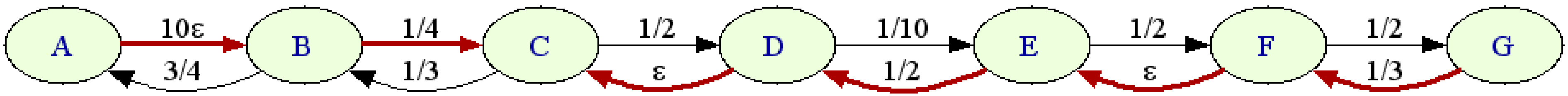
- An s-tree, T_s, is a complete collection of disjoint paths from states $\neq s$ to s.
- The likelihood of an s-tree T_s, written $\ell(T_s)$, $=_{Def}$ the product of its edge probabilities.
- The likelihood of a state s, written $\ell(s)$, $=_{Def}$ sum of the likelihood of all s-trees.

Theorem (Freidlin & Wentzell, 1984). Let P be an irreducible finite Markov process. Then, for all states, the likelihood of that state is proportional to the stationary probability of that state.

Counting s-trees with Freidlin & Wentzell: example



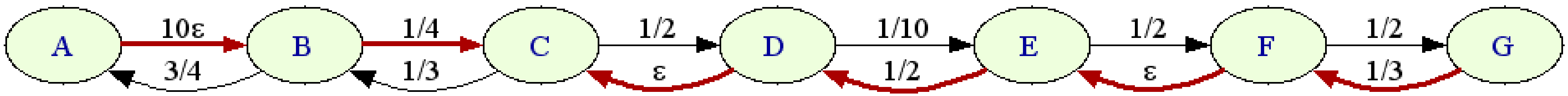
Counting s-trees with Freidlin & Wentzell: example



Freidlin & Wentzell (1984):

$$\mu(s) = \frac{v(s)}{\sum_{t \in S} v(t)}, \text{ where } \quad v(t) =_{Def} \sum_{T \in T_s} \ell(T_s)$$

Counting s-trees with Freidlin & Wentzell: example

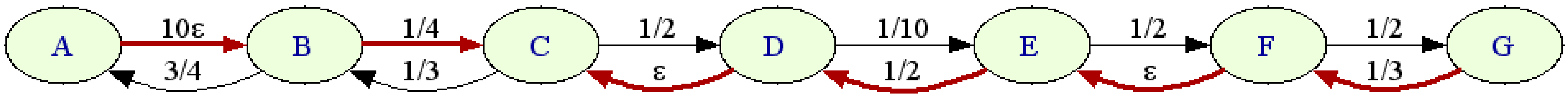


Freidlin & Wentzell (1984):

$$\mu(s) = \frac{v(s)}{\sum_{t \in S} v(t)}, \text{ where } \quad v(t) =_{Def} \sum_{T \in T_s} \ell(T_s)$$

The (unique) C-tree is coloured red. $\ell(T_C) = 10\epsilon \cdot 1/4 \cdot \cdots = 5\epsilon^3/12$.

Counting s-trees with Freidlin & Wentzell: example



Freidlin & Wentzell (1984):

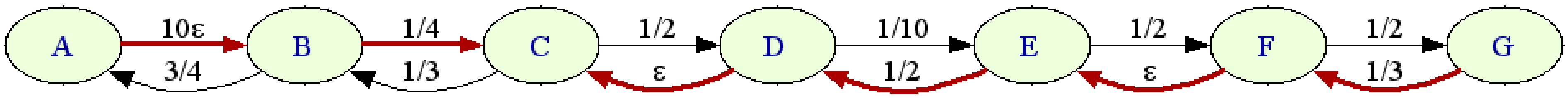
$$\mu(s) = \frac{v(s)}{\sum_{t \in S} v(t)}, \text{ where } \quad v(t) =_{Def} \sum_{T \in T_s} \ell(T_s)$$

The (unique) C-tree is coloured red. $\ell(T_C) = 10\epsilon \cdot 1/4 \cdot \dots = 5\epsilon^3/12$.

Similarly:

State:	A	B	C	D	E	F	G
Distribution:	$\epsilon^2/24$	$5\epsilon^3/9$	$5\epsilon^3/12$	$5\epsilon^2/24$	$\epsilon^2/24$	$\epsilon/48$	$\epsilon/32$

Counting s-trees with Freidlin & Wentzell: example



Freidlin & Wentzell (1984):

$$\mu(s) = \frac{v(s)}{\sum_{t \in S} v(t)}, \text{ where } \quad v(t) =_{Def} \sum_{T \in T_s} \ell(T_s)$$

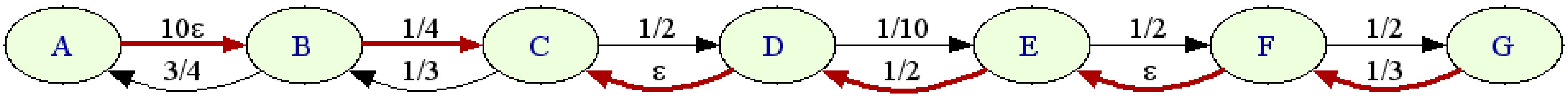
The (unique) C-tree is coloured red. $\ell(T_C) = 10\epsilon \cdot 1/4 \cdot \dots = 5\epsilon^3/12$.

Similarly:

State:	A	B	C	D	E	F	G
Distribution:	$\epsilon^2/24$	$5\epsilon^3/9$	$5\epsilon^3/12$	$5\epsilon^2/24$	$\epsilon^2/24$	$\epsilon/48$	$\epsilon/32$

If $\epsilon \to 0$, all ϵ-connections will be squeezed

Counting s-trees with Freidlin & Wentzell: example



Freidlin & Wentzell (1984):

$$\mu(s) = \frac{v(s)}{\sum_{t \in S} v(t)}, \text{ where } \quad v(t) =_{Def} \sum_{T \in T_s} \ell(T_s)$$

The (unique) C-tree is coloured red. $\ell(T_C) = 10\epsilon \cdot 1/4 \cdot \dots = 5\epsilon^3/12$.

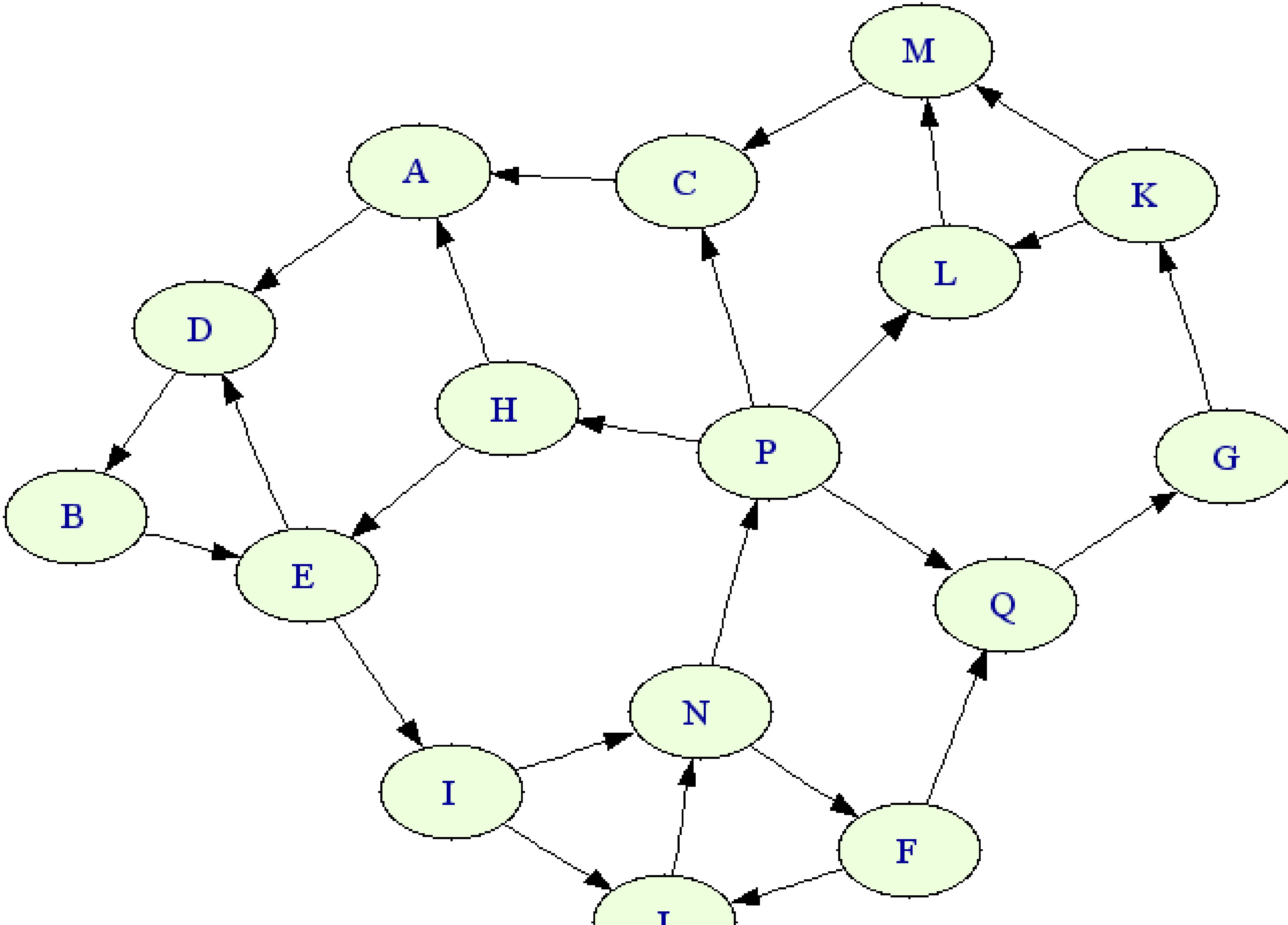
Similarly:

State:	A	B	C	D	E	F	G
Distribution:	$\epsilon^2/24$	$5\epsilon^3/9$	$5\epsilon^3/12$	$5\epsilon^2/24$	$\epsilon^2/24$	$\epsilon/48$	$\epsilon/32$

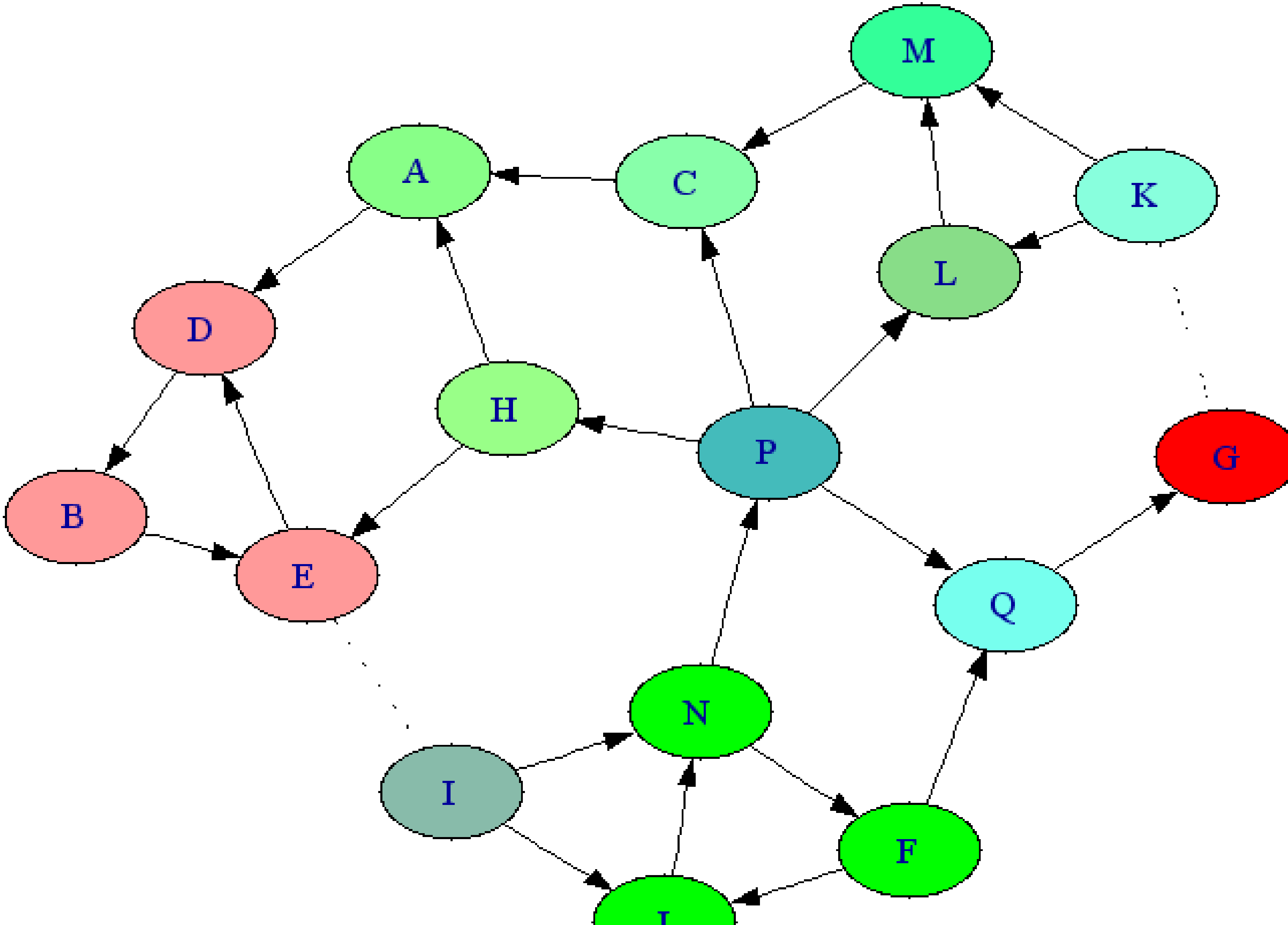
If $\epsilon \to 0$, all ϵ-connections will be squeezed, and F and G “win”.

Part 2:
Perturbed Markov processes

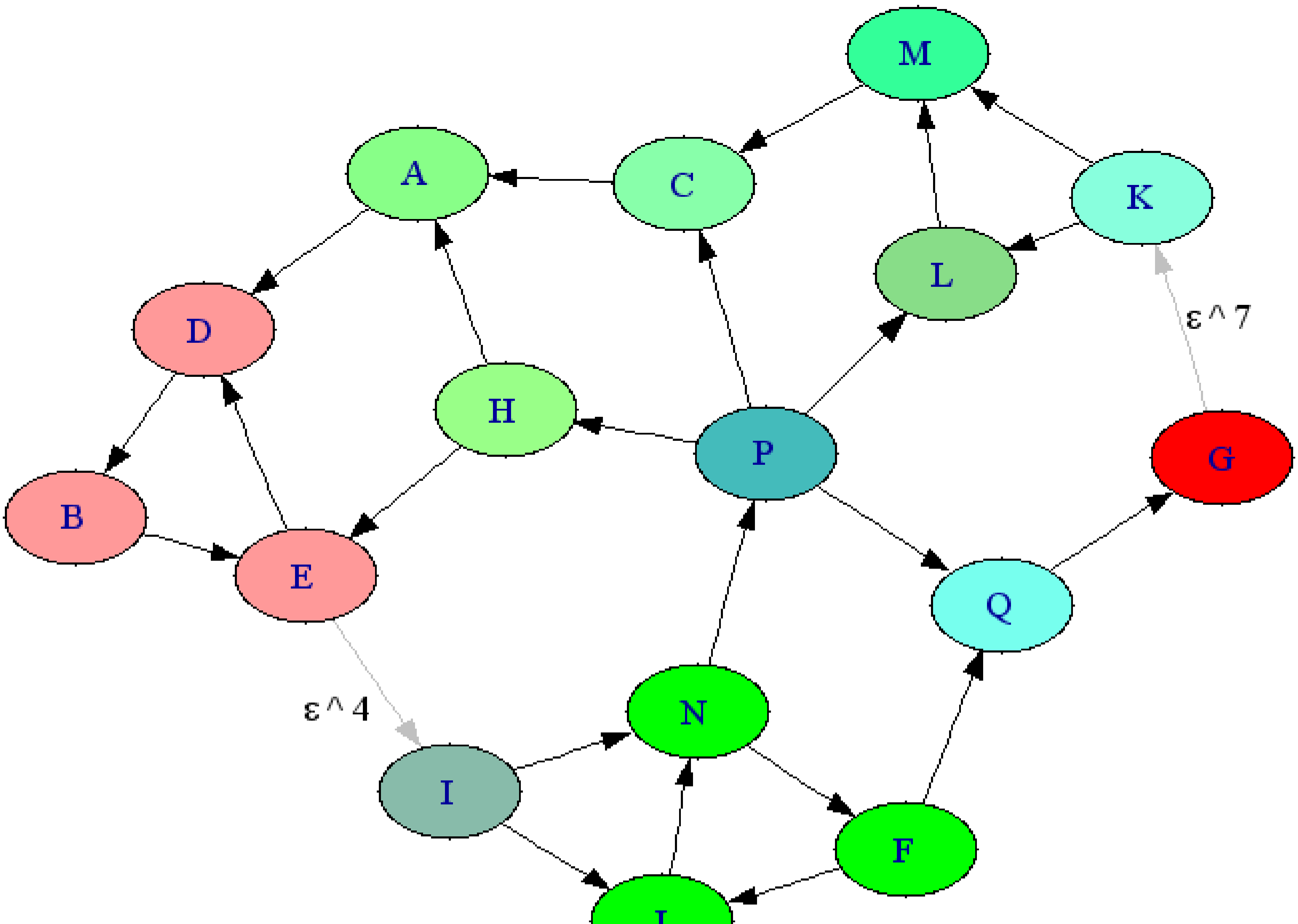
A Markov process may be ergodic (path-independent)



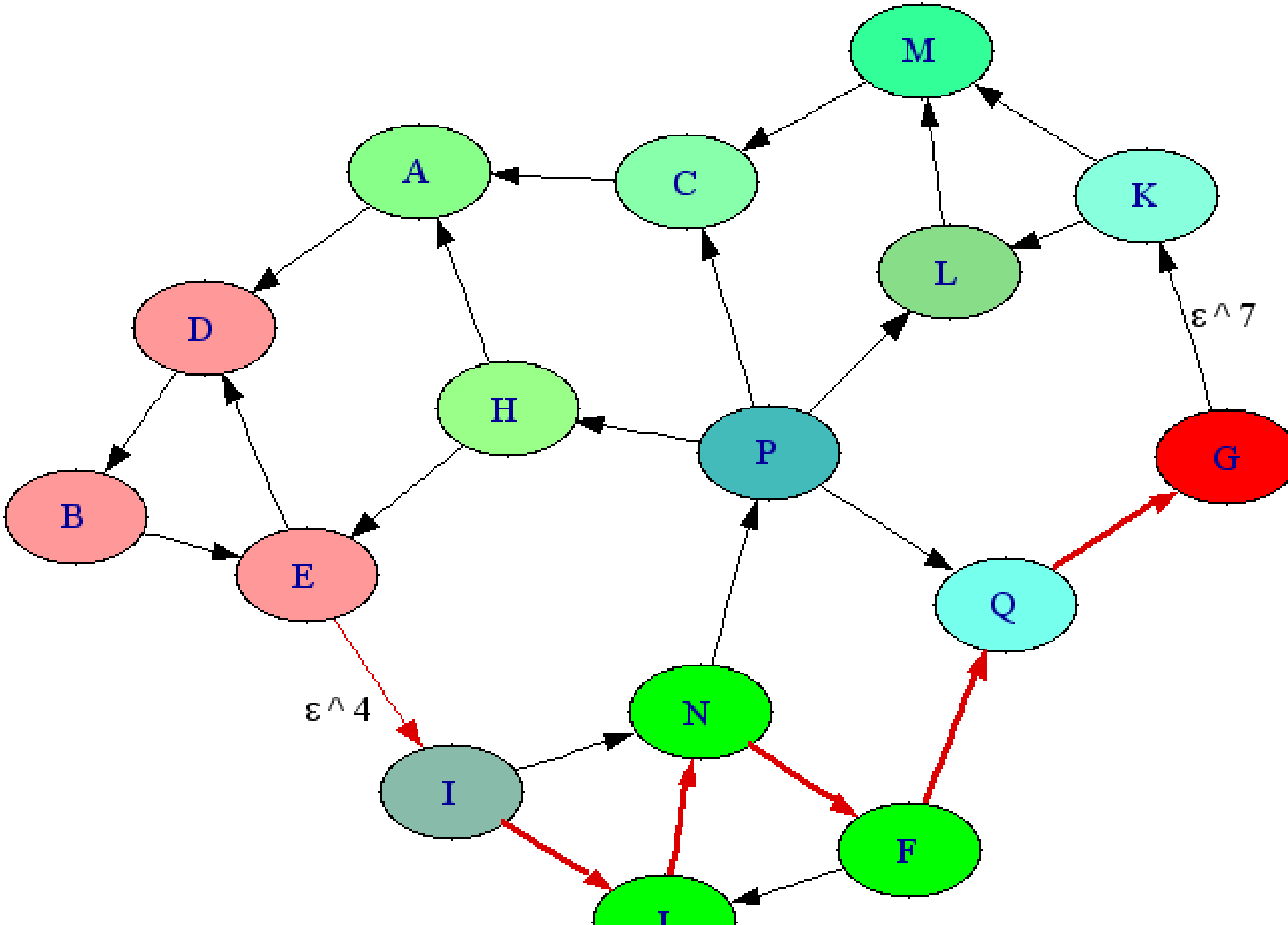
A Markov process may be non-ergodic (path-dependent)



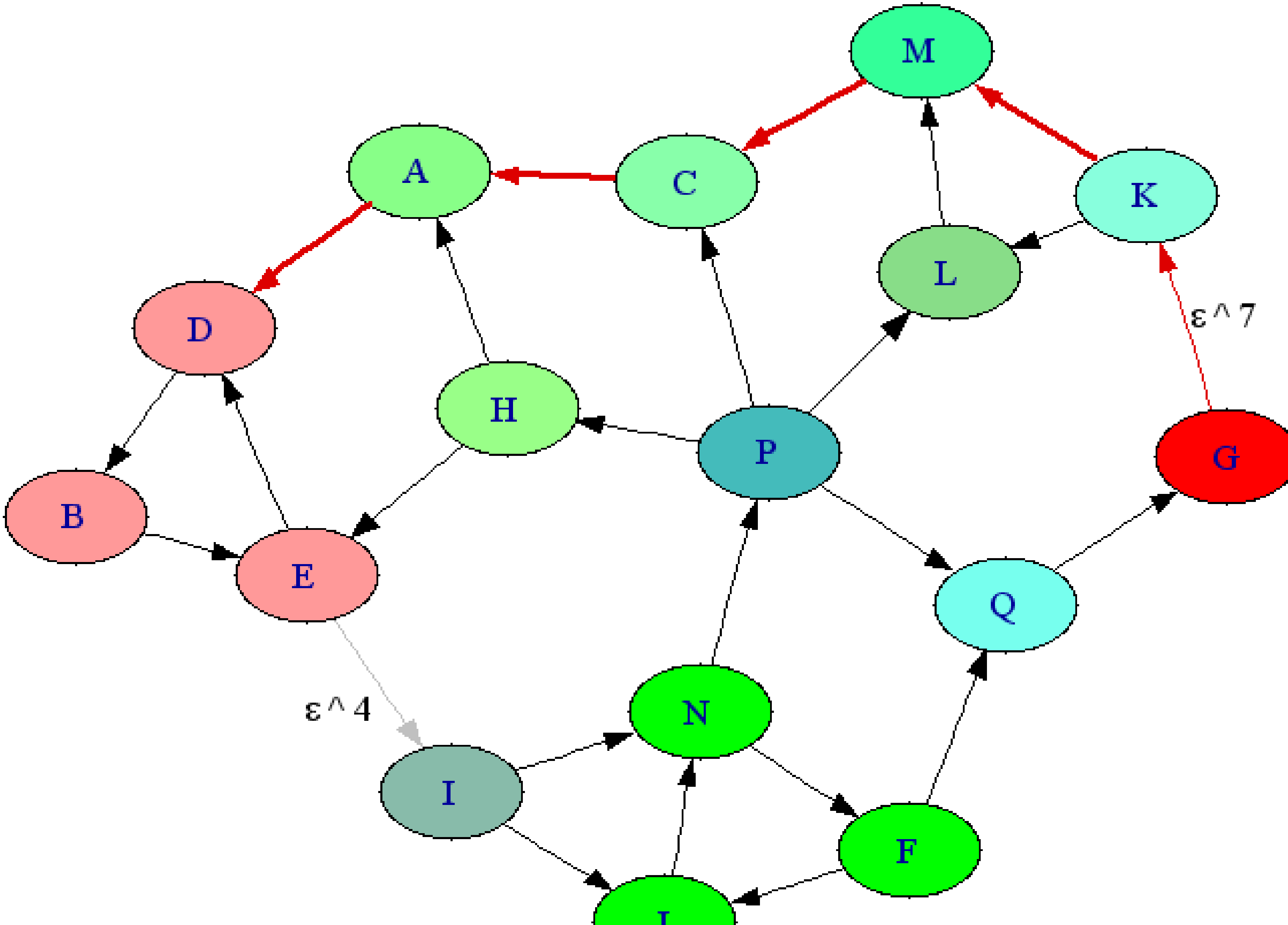
Make it ergodic by perturbing with $\epsilon^{r(s,s')}$



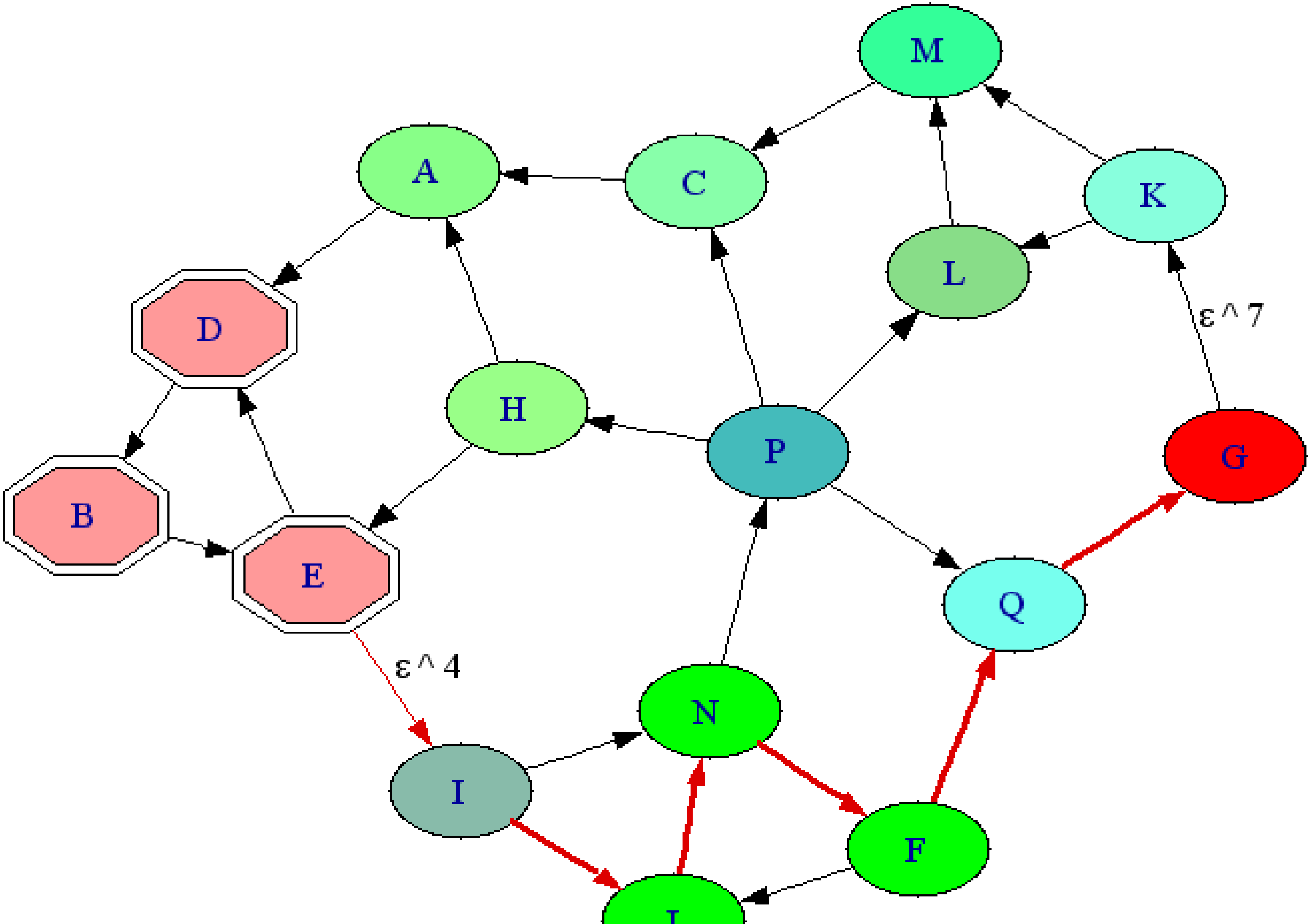
Compute s-trees from P^0-Recurrence classes only



Compute s-trees from P^0-Recurrence classes only



Class $\{B, D, E\}$ possesses lowest stochastic potential, viz.



Example of P^0 and P^ϵ

$$\lim_{\epsilon \to 0} \begin{pmatrix} 0.0 & 0.2 & 0.2 & 0.1 & 0.5 \\ 0.3 & \epsilon^7 & 0.1 & 0.1 & 0.5 - \epsilon^7 \\ 0.1 & 0.2 & 0.2 & 0.0 & 0.5 \\ 0.7 & 0.1 & 0.2 & 0.0 & 0.0 \\ 0.1 & 0.2 - \epsilon^2/2 & 0.2 & \epsilon^2 & 0.5 - \epsilon^2/2 \\ 0.0 & 0.0 & 0.1 & 0.0 & 0.9 \end{pmatrix}$$

$$= \begin{pmatrix} 0.0 & 0.2 & 0.2 & 0.1 & 0.5 \\ 0.3 & 0.0 & 0.1 & 0.1 & 0.5 \\ 0.1 & 0.2 & 0.2 & 0.0 & 0.5 \\ 0.7 & 0.1 & 0.2 & 0.0 & 0.0 \\ 0.1 & 0.2 & 0.2 & 0.0 & 0.5 \\ 0.0 & 0.0 & 0.1 & 0.0 & 0.9 \end{pmatrix}$$

Notice that some P^0-positive probabilities "have to give way" because row probabilities must add up to 1.

Perturbed Markov processes

Perturbed Markov processes

- P^0 is a Markov process on a finite state space S.

Perturbed Markov processes

- P^0 is a Markov process on a finite state space S.
- Let, for each $\epsilon \in (0, \epsilon^*]$, P^ϵ be a Markov process on the same state space.

Perturbed Markov processes

- P^0 is a Markov process on a finite state space S.
- Let, for each $\epsilon \in (0, \epsilon^*]$, P^ϵ be a Markov process on the same state space.
- The collection
$$\{P^\epsilon \mid \epsilon \in (0, \epsilon^*]\}$$
is said to be a regular perturbation of P^0

Perturbed Markov processes

- P^0 is a Markov process on a finite state space S.
- Let, for each $\epsilon \in (0, \epsilon^*]$, P^ϵ be a Markov process on the same state space.
- The collection

$$\{P^\epsilon \mid \epsilon \in (0, \epsilon^*]\}$$

is said to be a regular perturbation of P^0 if

1. Each P^ϵ is ergodic.

Perturbed Markov processes

- P^0 is a Markov process on a finite state space S.
- Let, for each $\epsilon \in (0, \epsilon^*]$, P^ϵ be a Markov process on the same state space.
- The collection

$$\{P^\epsilon \mid \epsilon \in (0, \epsilon^*]\}$$

is said to be a regular perturbation of P^0 if

1. Each P^ϵ is ergodic.
2. It holds that $\lim_{\epsilon \to 0} P^\epsilon = P^0$.

Perturbed Markov processes

- P^0 is a Markov process on a finite state space S.
- Let, for each $\epsilon \in (0, \epsilon^*]$, P^ϵ be a Markov process on the same state space.
- The collection

$$\{P^\epsilon \mid \epsilon \in (0, \epsilon^*]\}$$

is said to be a regular perturbation of P^0 if

1. Each P^ϵ is ergodic.
2. It holds that $\lim_{\epsilon \to 0} P^\epsilon = P^0$.
3. If $P^\epsilon_{s,s'} > 0$, for some $\epsilon > 0$, then $0 < \lim_{\epsilon \to 0} \dfrac{P^\epsilon}{\epsilon^{r(s,s')}} < \infty$

   for some $r(s, s') \geq 0$.

Perturbed Markov processes

- P^0 is a Markov process on a finite state space S.
- Let, for each $\epsilon \in (0, \epsilon^*]$, P^ϵ be a Markov process on the same state space.
- The collection

$$\{P^\epsilon \mid \epsilon \in (0, \epsilon^*]\}$$

is said to be a regular perturbation of P^0 if

1. Each P^ϵ is ergodic.
2. It holds that $\lim_{\epsilon \to 0} P^\epsilon = P^0$.
3. If $P^\epsilon_{s,s'} > 0$, for some $\epsilon > 0$, then $0 < \lim_{\epsilon \to 0} \dfrac{P^\epsilon}{\epsilon^{r(s,s')}} < \infty$

   for some $r(s, s') \geq 0$. This number is called the resistance from s to s'.

Resistance

Resistance

1. Each P^ϵ is ergodic.

Resistance

1. Each P^{ϵ} is ergodic.
2. It holds that $\lim_{\epsilon \to 0} P^{\epsilon} = P^{0}$.

Resistance

1. Each P^{ϵ} is ergodic.
2. It holds that $\lim_{\epsilon \to 0} P^{\epsilon} = P^{0}$.
3. If $P^{\epsilon}_{s,s'} > 0$, for some $\epsilon > 0$, then

$$0 \;<\; \lim_{\epsilon \to 0} \frac{P^{\epsilon}}{\epsilon^{r(s,s')}} \;<\; \infty$$

   for some $r(s, s') \geq 0$.

Resistance

1. Each P^{ϵ} is ergodic.
2. It holds that $\lim_{\epsilon \to 0} P^{\epsilon} = P^{0}$.
3. If $P^{\epsilon}_{s,s'} > 0$, for some $\epsilon > 0$, then

   $$0 \; < \; \lim_{\epsilon \to 0} \frac{P^{\epsilon}}{\epsilon^{r(s,s')}} \; < \; \infty$$

   for some $r(s, s') \geq 0$.
4. For transitions $s \to s'$ where $P^{0}_{s,s'} = P^{\epsilon}_{s,s'} = 0$ the resistance is defined to be ∞.

Resistance

1. Each P^{ϵ} is ergodic.
2. It holds that $\lim_{\epsilon \to 0} P^{\epsilon} = P^{0}$.
3. If $P^{\epsilon}_{s,s'} > 0$, for some $\epsilon > 0$, then

   $$0 < \lim_{\epsilon \to 0} \frac{P^{\epsilon}}{\epsilon^{r(s,s')}} < \infty$$

   for some $r(s, s') \geq 0$.
4. For transitions $s \to s'$ where $P^{0}_{s,s'} = P^{\epsilon}_{s,s'} = 0$ the resistance is defined to be ∞.

Notice:

Resistance

1. Each P^ϵ is ergodic.
2. It holds that $\lim_{\epsilon \to 0} P^\epsilon = P^0$.
3. If $P^\epsilon_{s,s'} > 0$, for some $\epsilon > 0$, then

$$0 \; < \; \lim_{\epsilon \to 0} \frac{P^\epsilon}{\epsilon^{r(s,s')}} \; < \; \infty$$

   for some $r(s, s') \geq 0$.
4. For transitions $s \to s'$ where $P^0_{s,s'} = P^\epsilon_{s,s'} = 0$ the resistance is defined to be ∞.

Notice:

- The number $r(s, s')$ is well-defined.

Resistance

1. Each P^ϵ is ergodic.
2. It holds that $\lim_{\epsilon\to 0} P^\epsilon = P^0$.
3. If $P^\epsilon_{s,s'} > 0$, for some $\epsilon > 0$, then

$$0 \;<\; \lim_{\epsilon\to 0} \frac{P^\epsilon}{\epsilon^{r(s,s')}} \;<\; \infty$$

   for some $r(s,s') \geq 0$.
4. For transitions $s \to s'$ where $P^0_{s,s'} = P^\epsilon_{s,s'} = 0$ the resistance is defined to be ∞.

Notice:

- The number $r(s,s')$ is well-defined.
- If $P^0_{s,s'} > 0$ then $r(s,s') = 0$.

Resistance

1. Each P^{ϵ} is ergodic.
2. It holds that $\lim_{\epsilon \to 0} P^{\epsilon} = P^0$.
3. If $P^{\epsilon}_{s,s'} > 0$, for some $\epsilon > 0$, then

$$0 < \lim_{\epsilon \to 0} \frac{P^{\epsilon}}{\epsilon^{r(s,s')}} < \infty$$

   for some $r(s, s') \geq 0$.
4. For transitions $s \to s'$ where $P^0_{s,s'} = P^{\epsilon}_{s,s'} = 0$ the resistance is defined to be ∞.

Notice:

- The number $r(s, s')$ is well-defined.
- If $P^0_{s,s'} > 0$ then $r(s, s') = 0$.
- If $r(s, s') = 0$ then $P^0_{s,s'} > 0$.

Stochastic stability

Stochastic stability

- Because each P^{ϵ} is ergodic, the stationary distribution μ^{ϵ} is uniquely determined, for every $\epsilon \in (0, \epsilon *]$.

Stochastic stability

- Because each P^ϵ is ergodic, the stationary distribution μ^ϵ is uniquely determined, for every $\epsilon \in (0, \epsilon*]$.
- It can be shown that

Stochastic stability

- Because each P^{ϵ} is ergodic, the stationary distribution μ^{ϵ} is uniquely determined, for every $\epsilon \in (0, \epsilon*]$.
- It can be shown that
  - $\mu^{0} =_{Def} \lim_{\epsilon \to 0} \mu^{\epsilon}(s)$ exists $\forall s$.

Stochastic stability

- Because each P^{ϵ} is ergodic, the stationary distribution μ^{ϵ} is uniquely determined, for every $\epsilon \in (0, \epsilon *]$.
- It can be shown that
  - $\mu^0 =_{Def} \lim_{\epsilon \to 0} \mu^{\epsilon}(s)$ exists $\forall s$.
  - μ^0 is a distribution on P^0.

Stochastic stability

- Because each P^ϵ is ergodic, the stationary distribution μ^ϵ is uniquely determined, for every $\epsilon \in (0, \epsilon*]$.
- It can be shown that
  - $\mu^0 =_{Def} \lim_{\epsilon \to 0} \mu^\epsilon(s)$ exists $\forall s$.
  - μ^0 is a distribution on P^0.
  - μ^0 is stationary for P^0.

Stochastic stability

- Because each P^{ϵ} is ergodic, the stationary distribution μ^{ϵ} is uniquely determined, for every $\epsilon \in (0, \epsilon*]$.
- It can be shown that
  - $\mu^0 =_{Def} \lim_{\epsilon \to 0} \mu^{\epsilon}(s)$ exists $\forall s$.
  - μ^0 is a distribution on P^0.
  - μ^0 is stationary for P^0.

Definition. A state s is said to be stochastically stable if

$$\mu^0(s) > 0$$

A way to compute stochastically stable states

A way to compute stochastically stable states

- Recurrence classes $1, \ldots, K$.

A way to compute stochastically stable states

- Recurrence classes $1, \ldots, K$.
- The resistance of a path from class i to class $j =_{Def}$ the **sum** of edge resistances.

A way to compute stochastically stable states

- Recurrence classes $1, \dots, K$.
- The resistance of a path from class i to class $j =_{Def}$ the **sum** of edge resistances.
- Construct edges r_{ij} (between classes) with the minimum resistance from i to j.

A way to compute stochastically stable states

- Recurrence classes $1, \dots, K$.
- The resistance of a path from class i to class $j =_{Def}$ the **sum** of edge resistances.
- Construct edges r_{ij} (between classes) with the minimum resistance from i to j.
- The resistance of j-tree T_j, written $r(T_j)$, $=_{Def}$ sum of edge resistances (in the class graph).

A way to compute stochastically stable states

- Recurrence classes $1, \ldots, K$.
- The resistance of a path from class i to class $j =_{Def}$ the **sum** of edge resistances.
- Construct edges r_{ij} (between classes) with the minimum resistance from i to j.
- The resistance of j-tree T_j, written $r(T_j)$, $=_{Def}$ sum of edge resistances (in the class graph).
- The stochastic potential of class j, written $p(j)$, $=_{Def}$ the minimum resistance over all j-trees.

A way to compute stochastically stable states

- Recurrence classes $1, \ldots, K$.
- The resistance of a path from class i to class $j =_{Def}$ the **sum** of edge resistances.
- Construct edges r_{ij} (between classes) with the minimum resistance from i to j.
- The resistance of j-tree T_j, written $r(T_j)$, $=_{Def}$ sum of edge resistances (in the class graph).
- The stochastic potential of class j, written $p(j)$, $=_{Def}$ the minimum resistance over all j-trees.

Theorem (Young, 1993). Let

$$\{ P^{\epsilon} \mid \epsilon \in (0, \epsilon^*] \}$$

be a regular perturbed Markov process.

A way to compute stochastically stable states

- Recurrence classes $1, \ldots, K$.
- The resistance of a path from class i to class $j =_{Def}$ the **sum** of edge resistances.
- Construct edges r_{ij} (between classes) with the minimum resistance from i to j.
- The resistance of j-tree T_j, written $r(T_j)$, $=_{Def}$ sum of edge resistances (in the class graph).
- The stochastic potential of class j, written $p(j)$, $=_{Def}$ the minimum resistance over all j-trees.

Theorem (Young, 1993). Let

$$\{ P^{\epsilon} \mid \epsilon \in (0, \epsilon^*] \}$$

be a regular perturbed Markov process.

Then the stochastically stable states

$$S^* = \{s \in S \mid \mu^0(s) > 0\}$$

are precisely those that are contained in the Recurrence class(es) of P^0 with minimum stochastic potential.

Minimum path resistance: example

Minimum path resistance: example

- Compute path resistance between all K Recurrence classes.

Minimum path resistance: example

- Compute path resistance between all K Recurrence classes.
- With K Recurrence classes there are always $K(K-1)$ minimum path resistances to be computed.

Example:

Minimum path resistance: example

- Compute path resistance between all K Recurrence classes.
- With K Recurrence classes there are always $K(K-1)$ minimum path resistances to be computed.

Example:

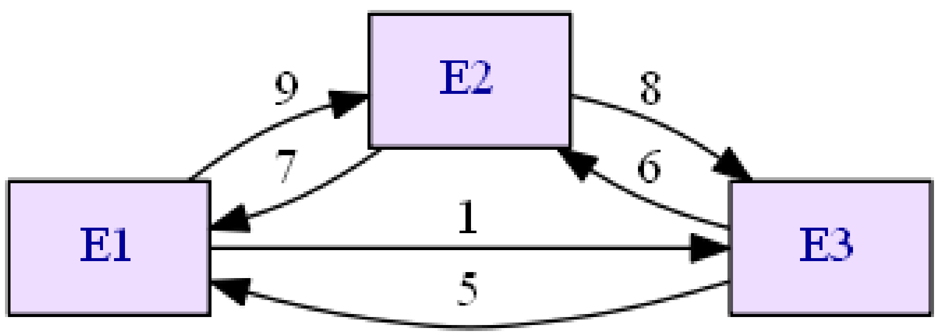
- Suppose there are three Recurrence classes E_1, E_2, and E_3.

Minimum path resistance: example

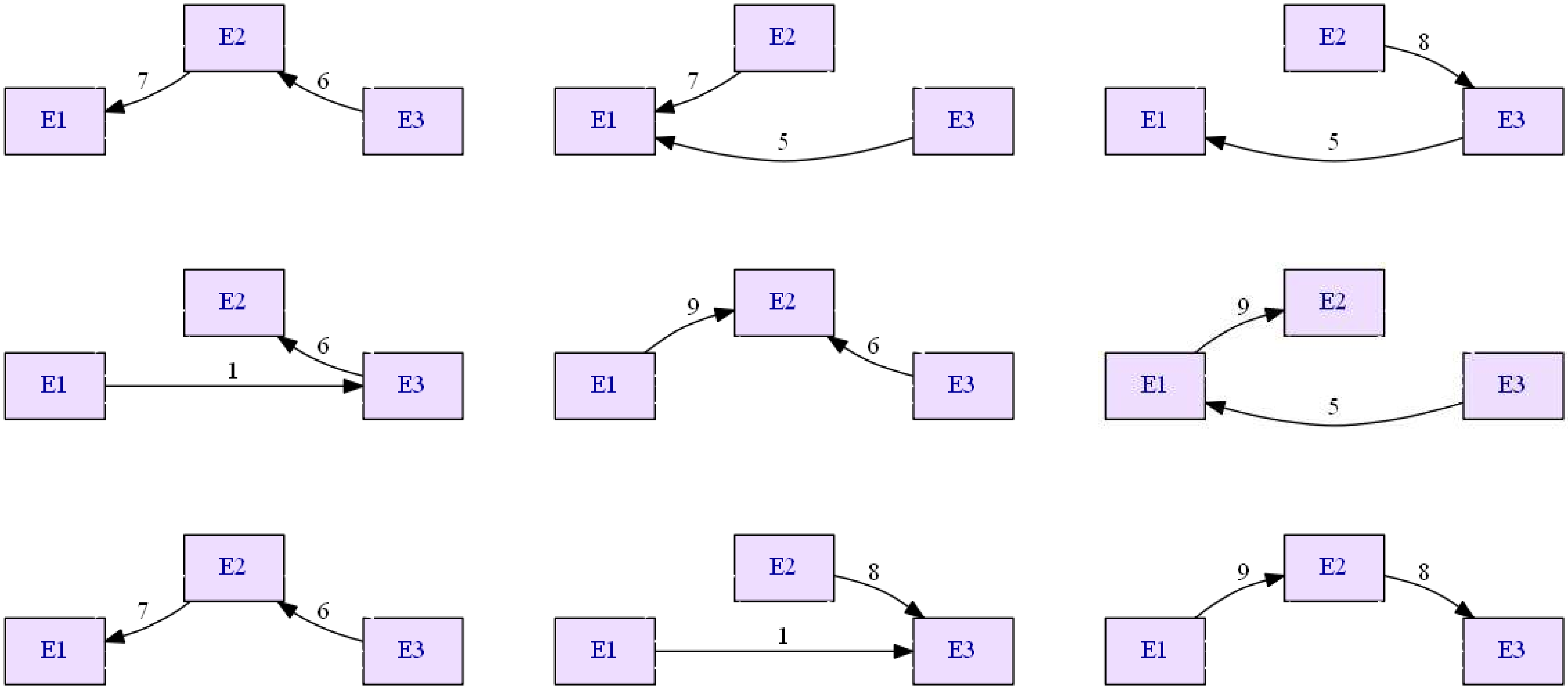
- Compute path resistance between all K Recurrence classes.
- With K Recurrence classes there are always $K(K-1)$ minimum path resistances to be computed.

Example:

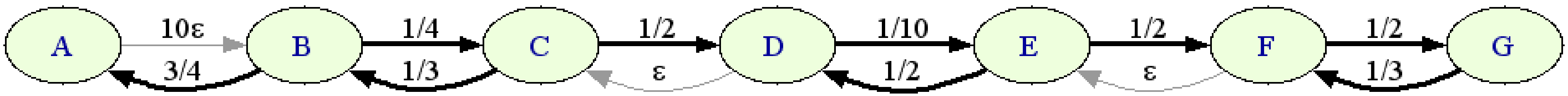
- Suppose there are three Recurrence classes E_1, E_2, and E_3.
- Minimum path resistances here are 1, 5, 6, 7, 8, 9.



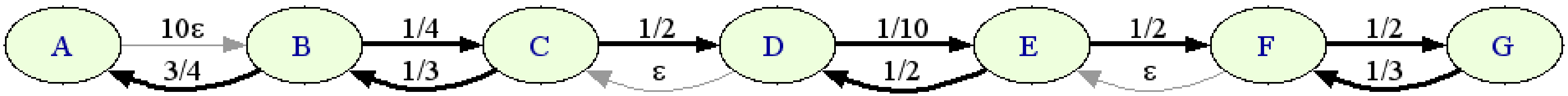
Nine j-trees generated by three recurrence classes



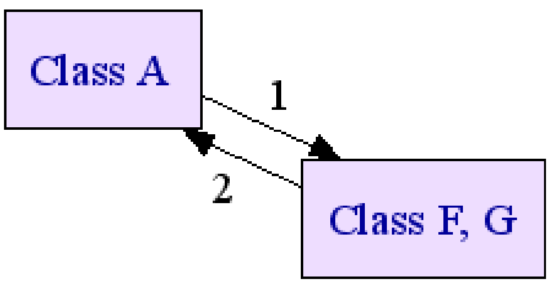
Revisit earlier example



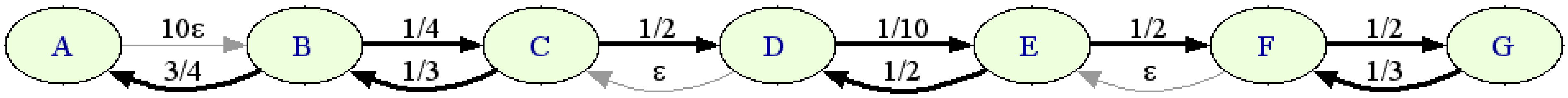
Revisit earlier example



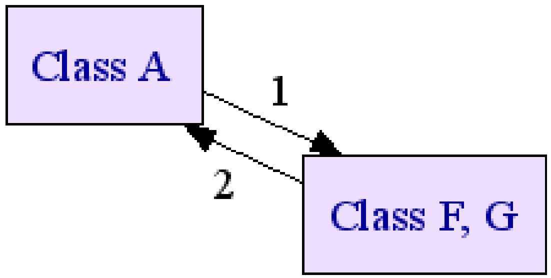
1. The unperturbed Markov process P^0 possesses two Recurrence classes, viz. $E_1 = \{A\}$ and $E_2 = \{F, G\}$.



Revisit earlier example

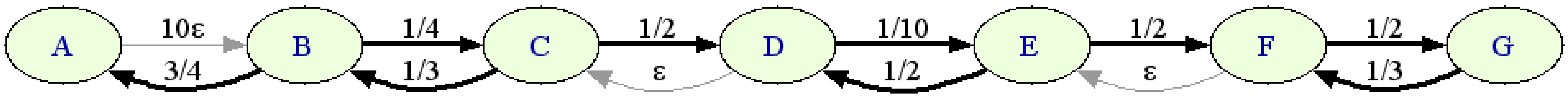


1. The unperturbed Markov process P^0 possesses two Recurrence classes, viz. $E_1 = \{A\}$ and $E_2 = \{F, G\}$.

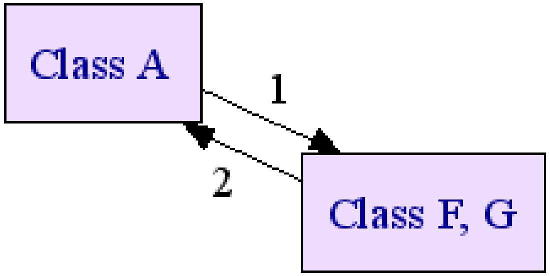


2. Least resistance from E_1 to E_2 is $10\epsilon \cdot \ldots = \epsilon^1/32$. Resistance 1.

Revisit earlier example



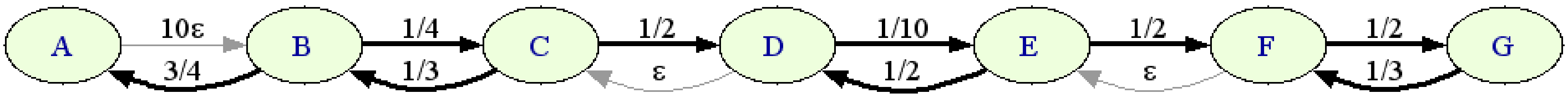
1. The unperturbed Markov process P^0 possesses two Recurrence classes, viz. $E_1 = \{A\}$ and $E_2 = \{F, G\}$.



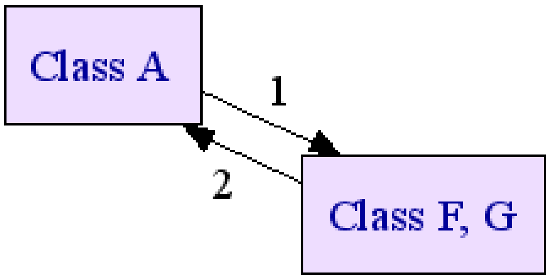
2. Least resistance from E_1 to E_2 is $10\epsilon \cdot \ldots = \epsilon^1/32$. Resistance 1.

3. Least resistance from E_2 to E_1 is $1/3 \cdot \epsilon \cdot \ldots = \epsilon^2/24$. Resistance 2.

Revisit earlier example



1. The unperturbed Markov process P^0 possesses two Recurrence classes, viz. $E_1 = \{A\}$ and $E_2 = \{F, G\}$.

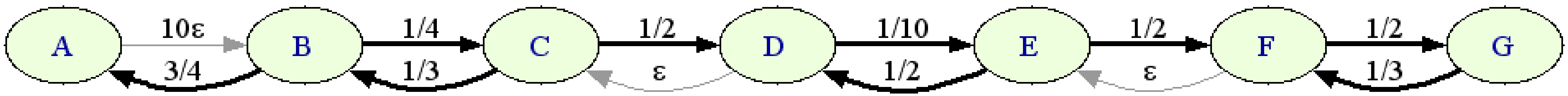


2. Least resistance from E_1 to E_2 is $10\epsilon \cdot \ldots = \epsilon^1/32$. Resistance 1.

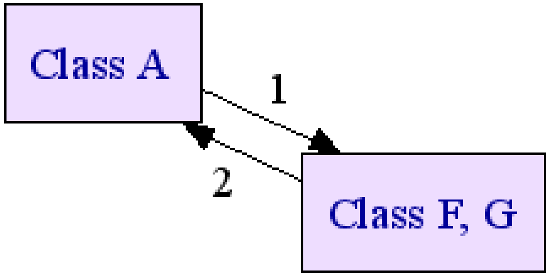
3. Least resistance from E_2 to E_1 is $1/3 \cdot \epsilon \cdot \ldots = \epsilon^2/24$. Resistance 2.

4. There is only one resistance tree to either side, hence one minimum resistance tree.

Revisit earlier example



1. The unperturbed Markov process P^0 possesses two Recurrence classes, viz. $E_1 = \{A\}$ and $E_2 = \{F, G\}$.



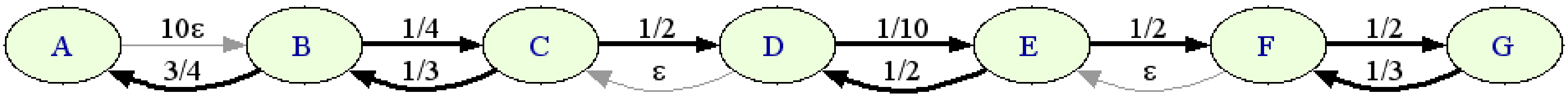
2. Least resistance from E_1 to E_2 is $10\epsilon \cdot \ldots = \epsilon^1/32$. Resistance 1.

3. Least resistance from E_2 to E_1 is $1/3 \cdot \epsilon \cdot \ldots = \epsilon^2/24$. Resistance 2.

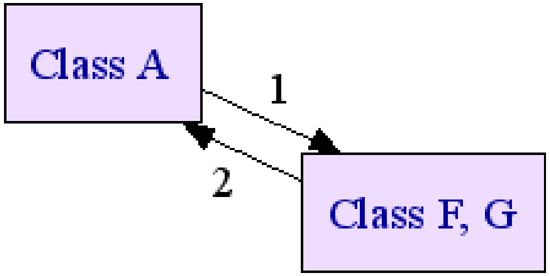
4. There is only one resistance tree to either side, hence one minimum resistance tree.

5. Stochastic potential of E_1 is 2; stochastic potential of E_2 is 1.

Revisit earlier example



1. The unperturbed Markov process P^0 possesses two Recurrence classes, viz. $E_1 = \{A\}$ and $E_2 = \{F, G\}$.



2. Least resistance from E_1 to E_2 is $10\epsilon \cdot \ldots = \epsilon^1/32$. Resistance 1.

3. Least resistance from E_2 to E_1 is $1/3 \cdot \epsilon \cdot \ldots = \epsilon^2/24$. Resistance 2.

4. There is only one resistance tree to either side, hence one minimum resistance tree.

5. Stochastic potential of E_1 is 2; stochastic potential of E_2 is 1.

6. Conclusion: E_2 is stochastically stable, E_1 is not.

Part 3: Applications

Idiosyncratic play in technology adoption

"How, then, might institutional change occur? Because best-response play renders both conventions absorbing states, it is clear that in order to understand institutional change, some kind of nonbest-response play must be introduced. Suppose there is a probability ϵ that when individuals are in the process of updating, each may switch their type for idiosyncratic reasons. Thus, $1 - \epsilon$ represents the probability that the individual pursues the best-response updating process described above. The idiosyncratic play accounting for nonbest responses need not be irrational or odd; it simply represents actions whose reasons are not explicitly modeled. Included is experimentation, whim, error, and intentional acts seeking to affect game outcomes but whose motivations are not captured by the above game."

From *Microeconomics: behavior, institutions, and evolution* (Bowles, 2003).

Technology adoption

		Other:	
		Operating system A	Operating system B
You:	Operating system A	(a, a)	$(0, 0)$
	Operating system B	$(0, 0)$	(b, b)

Technology adoption

		Other:	
		Operating system A	Operating system B
You:	Operating system A	(a, a)	$(0, 0)$
	Operating system B	$(0, 0)$	(b, b)

Total number of players : N, for example $N = 5$

Technology adoption

		Other: Operating system A	Operating system B
You:	Operating system A	(a, a)	$(0, 0)$
	Operating system B	$(0, 0)$	(b, b)

$$\text{Total number of players} : N, \text{ for example } N = 5$$
$$\text{Total number of players } A : K, \text{ for example } K = 3$$

Technology adoption

		Other:	
		Operating system A	Operating system B
You:	Operating system A	(a, a)	$(0, 0)$
	Operating system B	$(0, 0)$	(b, b)

$$\text{Total number of players} : N, \text{ for example } N = 5$$
$$\text{Total number of players } A : K, \text{ for example } K = 3$$
$$\text{Sample size} : n, \text{ for example } n = 4$$

Technology adoption

		Other: Operating system A	Operating system B
You:	Operating system A	(a, a)	$(0, 0)$
	Operating system B	$(0, 0)$	(b, b)

$$\begin{aligned}\text{Total number of players} &: N, \text{ for example } N = 5\\ \text{Total number of players } A &: K, \text{ for example } K = 3\\ \text{Sample size} &: n, \text{ for example } n = 4\\ \text{Number in sample playing } A &: k, \text{ for example } k = 2.\end{aligned}$$

Technology adoption

		Other: Operating system A	Operating system B
You:	Operating system A	(a, a)	$(0, 0)$
	Operating system B	$(0, 0)$	(b, b)

$$\begin{aligned}
\text{Total number of players} : N&, \text{ for example } N = 5 \\
\text{Total number of players } A : K&, \text{ for example } K = 3 \\
\text{Sample size} : n&, \text{ for example } n = 4 \\
\text{Number in sample playing } A : k&, \text{ for example } k = 2.
\end{aligned}$$

$$\begin{aligned}
P(\ \#A's = 2 \mid AABBB\ ) &= \text{hypergeometric}(N, K, n, 2) \\
&= \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \frac{\binom{3}{2}\binom{5-3}{4-2}}{\binom{5}{4}} = \frac{\binom{3}{2}\binom{2}{2}}{\binom{5}{4}} = \frac{3 \cdot 1}{5} - = \frac{3}{5}.
\end{aligned}$$

Technology adoption

		Other: Operating system A	Operating system B
You:	Operating system A	(a, a)	$(0, 0)$
	Operating system B	$(0, 0)$	(b, b)

$$\begin{aligned} \text{Total number of players} &: N, \text{ for example } N = 5 \\ \text{Total number of players } A &: K, \text{ for example } K = 3 \\ \text{Sample size} &: n, \text{ for example } n = 4 \\ \text{Number in sample playing } A &: k, \text{ for example } k = 2. \end{aligned}$$

$$\begin{aligned} P(\ \#A's = 2 \mid AABBB\ ) &= \text{hypergeometric}(N, K, n, 2) \\ &= \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \frac{\binom{3}{2}\binom{5-3}{4-2}}{\binom{5}{4}} = \frac{\binom{3}{2}\binom{2}{2}}{\binom{5}{4}} = \frac{3 \cdot 1}{5} - = \frac{3}{5}. \end{aligned}$$

This process is path-dependent (non-ergodic): for example always $BABBB, BABBB, \text{etc.} \to BBBBB$. With $b \gg a$ even $BAABB, \text{etc.} \to BBBBB$.

The tipping effect

$$\begin{aligned} \text{Total number of players} : N, &\quad \text{for example } N = 5 \\ \text{Sample size} : n, &\quad \text{for example } n = 3 \\ \text{Total number of players currently playing } A : K, &\quad \text{for example } K = 2 \end{aligned}$$

The tipping effect

$$\begin{aligned} \text{Total number of players} : N, &\quad \text{for example } N = 5 \\ \text{Sample size} : n, &\quad \text{for example } n = 3 \\ \text{Total number of players currently playing } A : K, &\quad \text{for example } K = 2 \end{aligned}$$

Suppose $a = b$.

The tipping effect

$$\begin{aligned}\text{Total number of players}: N, &\quad \text{for example } N=5\\ \text{Sample size}: n, &\quad \text{for example } n=3\\ \text{Total number of players currently playing } A: K, &\quad \text{for example } K=2\end{aligned}$$

Suppose $a = b$. Let

$$\begin{aligned}
E_1 &= \{\, s \in S \mid s \sim AAAAA \,\} &&= \{AAAAA\}\\
T_1 &= \{\, s \in S \mid s \sim AAAAB \,\} &&= \{AAAAB, AAABA, \dots, BAAAA\}\\
T_2 &= \{\, s \in S \mid s \sim AAABB \,\} &&= \{AAABB, AABAB, \dots, BBBAA\}\\
T_3 &= \{\, s \in S \mid s \sim ABBBB \,\}\\
E_2 &= \{\, s \in S \mid s \sim BBBBB \,\}
\end{aligned}$$

The tipping effect

$$\begin{aligned} \text{Total number of players} : N, &\quad \text{for example } N = 5 \\ \text{Sample size} : n, &\quad \text{for example } n = 3 \\ \text{Total number of players currently playing } A : K, &\quad \text{for example } K = 2 \end{aligned}$$

Suppose $a = b$. Let

$$\begin{aligned}
E_1 &= \{\, s \in S \mid s \sim AAAAA \,\} && = \{AAAAA\} \\
T_1 &= \{\, s \in S \mid s \sim AAAAB \,\} && = \{AAAAB, AAABA, \dots, BAAAA\} \\
T_2 &= \{\, s \in S \mid s \sim AAABB \,\} && = \{AAABB, AABAB, \dots, BBBAA\} \\
T_3 &= \{\, s \in S \mid s \sim ABBBB \,\} \\
E_2 &= \{\, s \in S \mid s \sim BBBBB \,\}
\end{aligned}$$

- How many idiosyncratic transitions must be made to move from E_1 to E_2?

The tipping effect

$$\begin{aligned} \text{Total number of players} : N, &\quad \text{for example } N = 5 \\ \text{Sample size} : n, &\quad \text{for example } n = 3 \\ \text{Total number of players currently playing } A : K, &\quad \text{for example } K = 2 \end{aligned}$$

Suppose $a = b$. Let

$$\begin{aligned}
E_1 &= \{\, s \in S \mid s \sim AAAAA \,\} && = \{AAAAA\} \\
T_1 &= \{\, s \in S \mid s \sim AAAAB \,\} && = \{AAAAB, AAABA, \dots, BAAAA\} \\
T_2 &= \{\, s \in S \mid s \sim AAABB \,\} && = \{AAABB, AABAB, \dots, BBBAA\} \\
T_3 &= \{\, s \in S \mid s \sim ABBBB \,\} \\
E_2 &= \{\, s \in S \mid s \sim BBBBB \,\}
\end{aligned}$$

- How many idiosyncratic transitions must be made to move from E_1 to E_2? What is the resistance from E_1 to E_2? From E_1 to E_2?

The tipping effect

$$
\begin{aligned}
\text{Total number of players} : N, &\quad \text{for example } N = 5 \\
\text{Sample size} : n, &\quad \text{for example } n = 3 \\
\text{Total number of players currently playing } A : K, &\quad \text{for example } K = 2
\end{aligned}
$$

Suppose $a = b$. Let

$$
\begin{aligned}
E_1 &= \{\, s \in S \mid s \sim AAAAA \,\} && = \{AAAAA\} \\
T_1 &= \{\, s \in S \mid s \sim AAAAB \,\} && = \{AAAAB, AAABA, \dots, BAAAA\} \\
T_2 &= \{\, s \in S \mid s \sim AAABB \,\} && = \{AAABB, AABAB, \dots, BBBAA\} \\
T_3 &= \{\, s \in S \mid s \sim ABBBB \,\} \\
E_2 &= \{\, s \in S \mid s \sim BBBBB \,\}
\end{aligned}
$$

- How many idiosyncratic transitions must be made to move from E_1 to E_2? What is the resistance from E_1 to E_2? From E_1 to E_2?
- What is (are) the stochastically stable state(s)?

Tipping point (general case)

Tipping point (general case)

- Suppose $b > a$ and everyone uses A.

Tipping point (general case)

- Suppose $b > a$ and everyone uses A.
- Generally, an individual will choose for B when

$$bk \geq a(n-k) \quad \Leftrightarrow \quad k \geq \frac{a}{a+b}n.$$

(Breaking ties if "$\geq$" = "$=$".)

Tipping point (general case)

- Suppose $b > a$ and everyone uses A.
- Generally, an individual will choose for B when

  $$bk \geq a(n-k) \quad \Leftrightarrow \quad k \geq \frac{a}{a+b}n.$$

  (Breaking ties if "$\geq$" = "$=$".)
- So

  $$\lceil \frac{a}{a+b}n \rceil$$

  times an idiosyncratic choice must be made to move from All-A into the first transient class that, without further idiosyncrasies, leads to All-B.

Tipping point (general case)

- Suppose $b > a$ and everyone uses A.
- Generally, an individual will choose for B when

  $$bk \geq a(n-k) \quad \Leftrightarrow \quad k \geq \frac{a}{a+b}n.$$

  (Breaking ties if "$\geq$" = "$=$".)
- So

  $$\lceil \frac{a}{a+b}n \rceil$$

  times an idiosyncratic choice must be made to move from All-A into the first transient class that, without further idiosyncrasies, leads to All-B.
- If $\Pr\{\text{idiosyncratic choice}\} = \epsilon$ then

  $$\Pr\{\text{tipping}\} = (\frac{\epsilon}{2})^{\lceil \frac{a}{a+b}n \rceil}.$$

Tipping point (general case)

- Suppose $b > a$ and everyone uses A.
- Generally, an individual will choose for B when

$$bk \geq a(n-k) \quad \Leftrightarrow \quad k \geq \frac{a}{a+b}n.$$

(Breaking ties if "$\geq$" = "$=$".)

- So

$$\lceil \frac{a}{a+b}n \rceil$$

times an idiosyncratic choice must be made to move from All-A into the first transient class that, without further idiosyncrasies, leads to All-B.

- If $\Pr\{\text{idiosyncratic choice}\} = \epsilon$ then

$$\Pr\{\text{tipping}\} = (\frac{\epsilon}{2})^{\lceil \frac{a}{a+b}n \rceil}.$$

Indeed $\epsilon/2$, if we assume that idiosyncrasy is uniformly distributed among A and B.

In that case, half of the idiosyncratic choices are contra-productive again!

Tipping point (general case)

- Suppose $b > a$ and everyone uses A.
- Generally, an individual will choose for B when

$$bk \geq a(n-k) \quad \Leftrightarrow \quad k \geq \frac{a}{a+b}n.$$

(Breaking ties if "$\geq$" = "$=$".)

- So

$$\lceil \frac{a}{a+b}n \rceil$$

times an idiosyncratic choice must be made to move from All-A into the first transient class that, without further idiosyncrasies, leads to All-B.

- If $\Pr\{\text{idiosyncratic choice}\} = \epsilon$ then

$$\Pr\{\text{tipping}\} = (\frac{\epsilon}{2})^{\lceil \frac{a}{a+b}n \rceil}.$$

Indeed $\epsilon/2$, if we assume that idiosyncrasy is uniformly distributed among A and B.

In that case, half of the idiosyncratic choices are contra-productive again!

- Because $b > a$, $r(\text{All-}B \rightarrow \text{All-}A) > r(\text{All-}A \rightarrow \text{All-}B)$, and all-$B$ has minimum stochastic potential

Tipping point (general case)

- Suppose $b > a$ and everyone uses A.
- Generally, an individual will choose for B when

$$bk \geq a(n-k) \quad \Leftrightarrow \quad k \geq \frac{a}{a+b}n.$$

(Breaking ties if "$\geq$" = "$=$".)

- So

$$\lceil \frac{a}{a+b}n \rceil$$

times an idiosyncratic choice must be made to move from All-A into the first transient class that, without further idiosyncrasies, leads to All-B.

- If $\Pr\{\text{idiosyncratic choice}\} = \epsilon$ then

$$\Pr\{\text{tipping}\} = (\frac{\epsilon}{2})^{\lceil \frac{a}{a+b}n \rceil}.$$

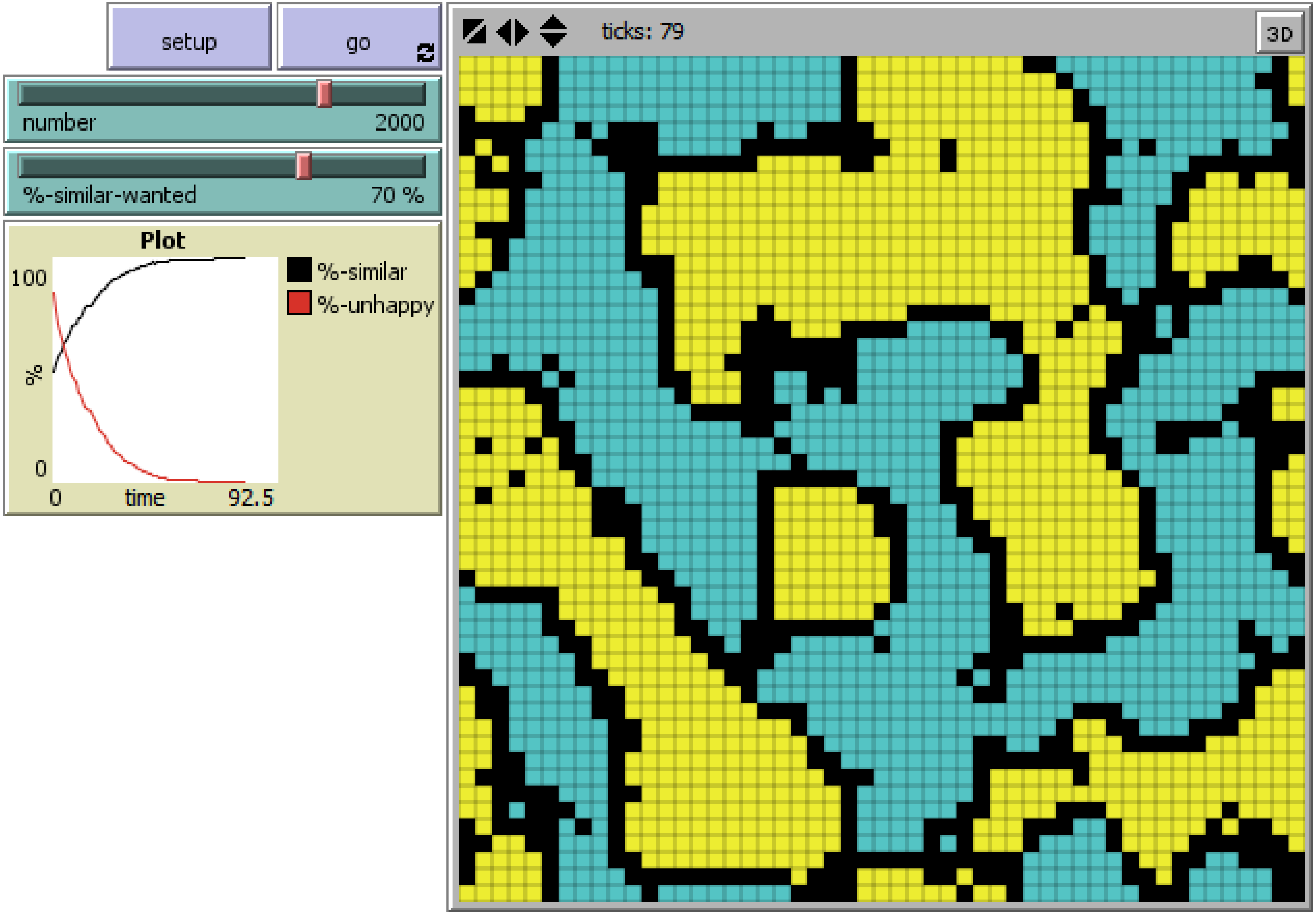
Indeed $\epsilon/2$, if we assume that idiosyncrasy is uniformly distributed among A and B.

In that case, half of the idiosyncratic choices are contra-productive again!

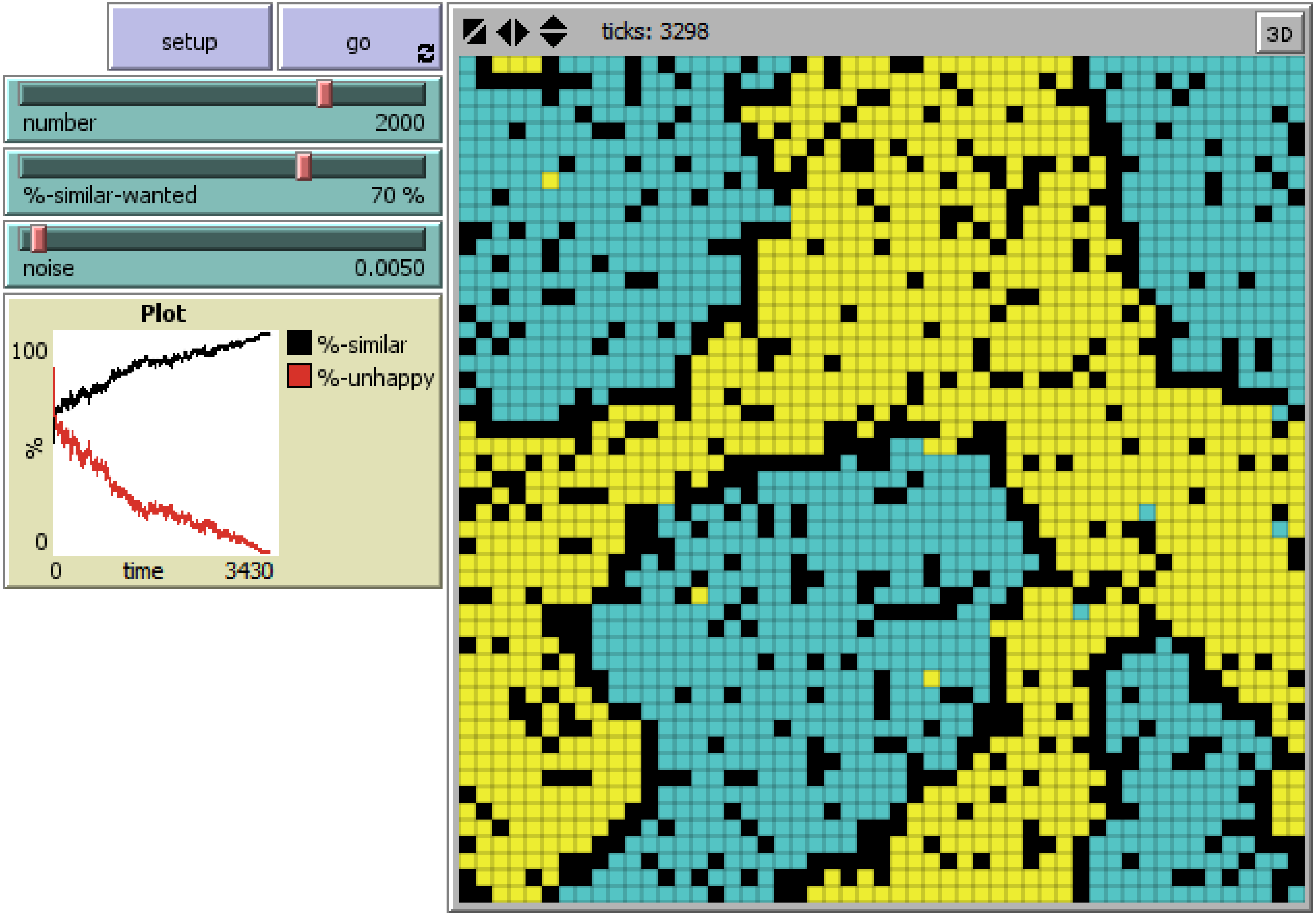
- Because $b > a$, $r(\text{All-}B \rightarrow \text{All-}A) > r(\text{All-}A \rightarrow \text{All-}B)$, and all-$B$ has minimum stochastic potential, and is therefore a Stoch. SS.

Part 4: Schelling's model of segregation

Schelling's model in 2D (torus)

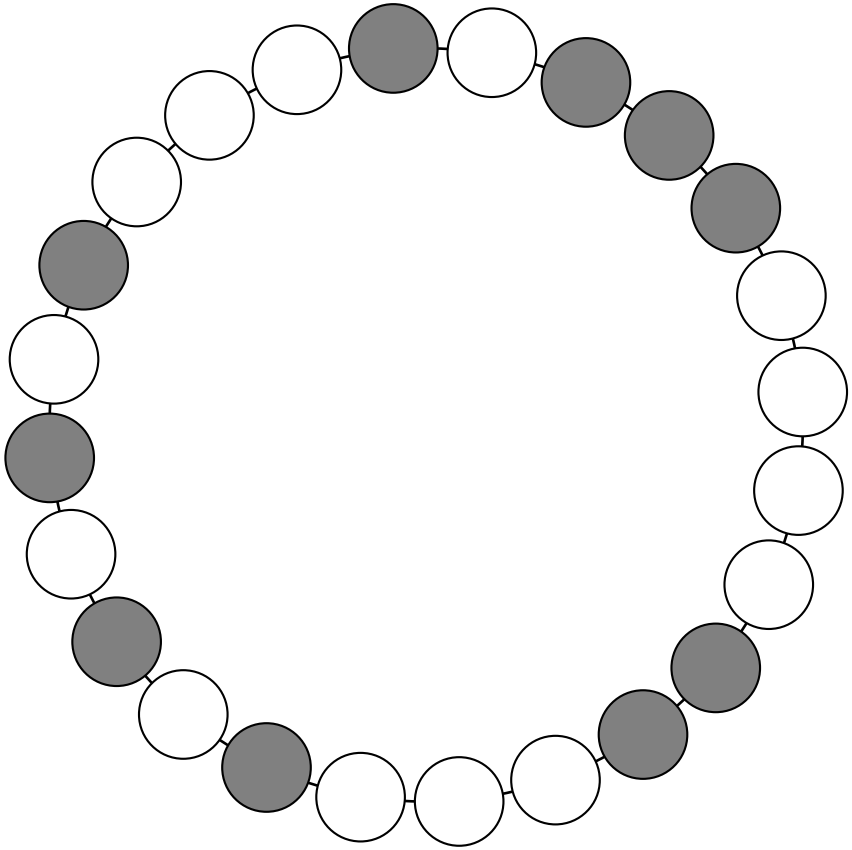


Schelling's model in 2D: P^ϵ for small ϵ

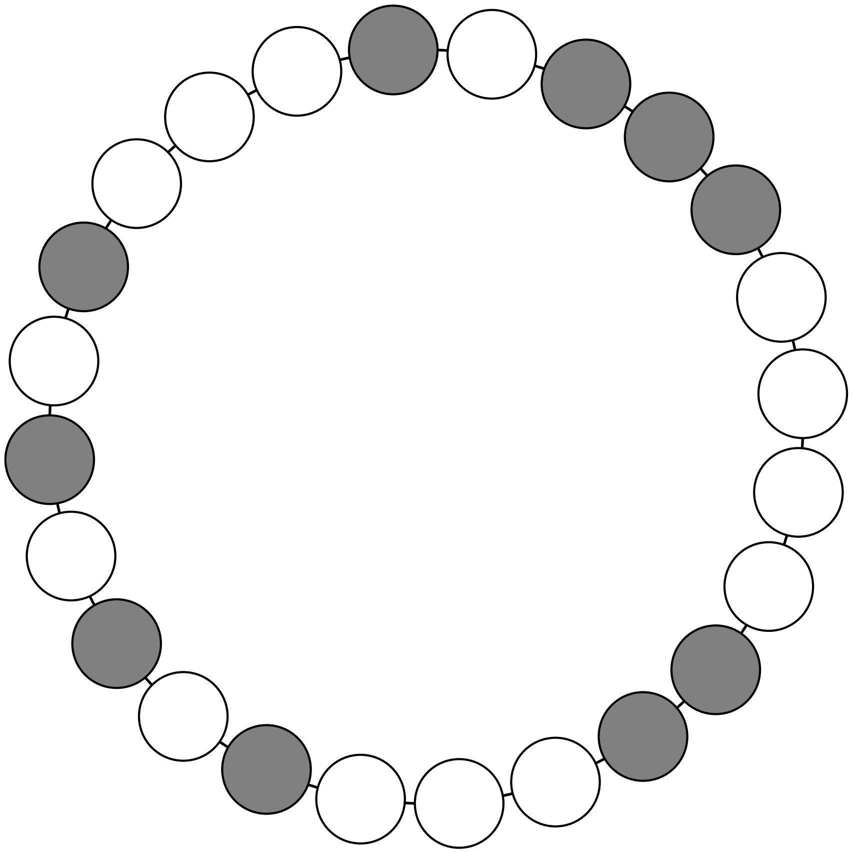


Schelling's model in 1D (circle)

Schelling's model in 1D (circle)

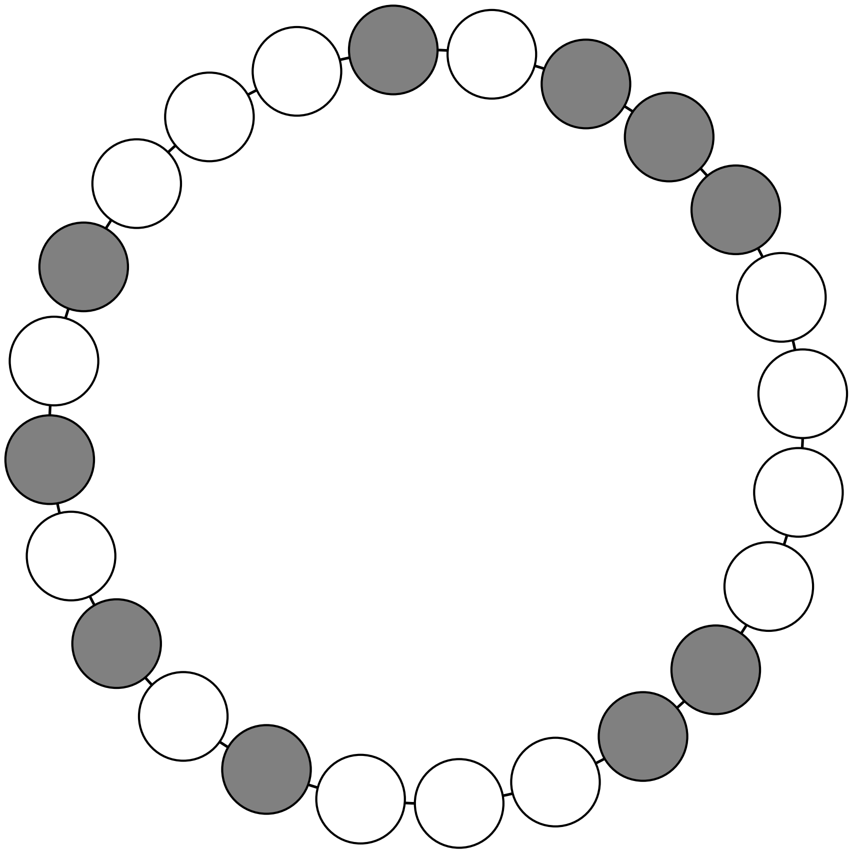


Schelling's model in 1D (circle)



- Schelling (1969, 1971, 1978).

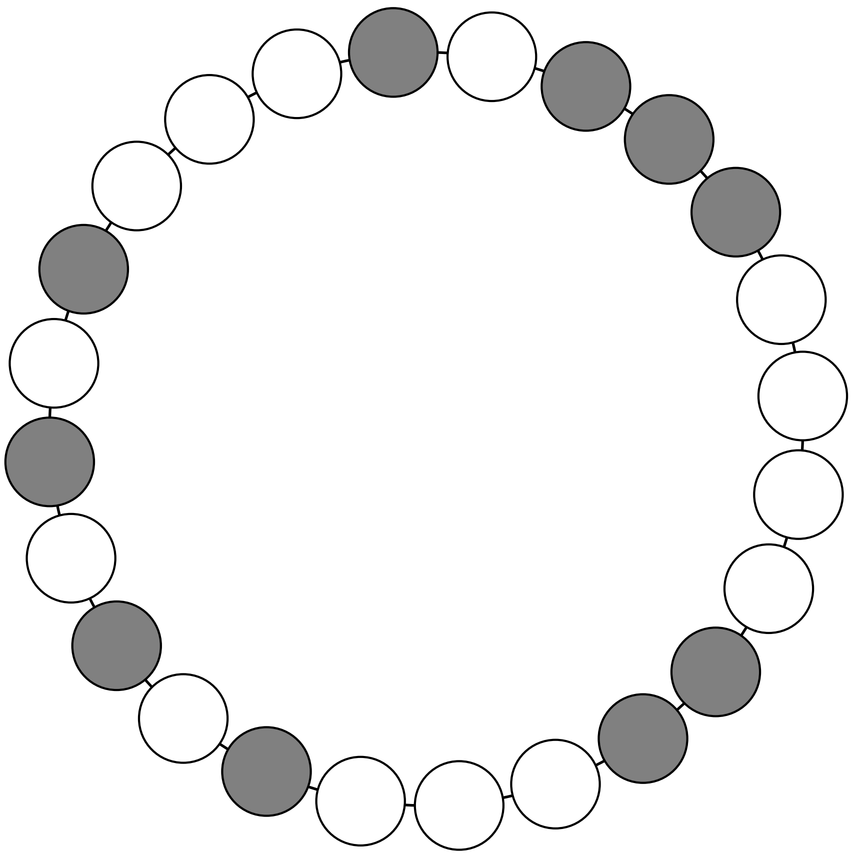
Schelling's model in 1D (circle)



other people are content.

- Schelling (1969, 1971, 1978).
- Isolated people are discontent,

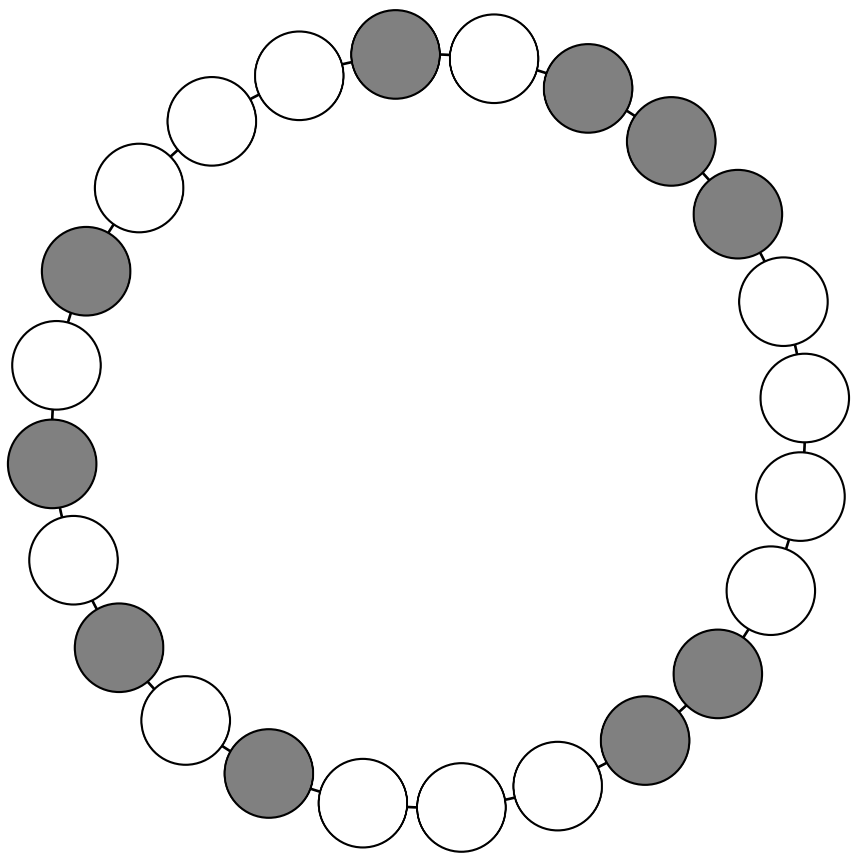
Schelling's model in 1D (circle)



- Schelling (1969, 1971, 1978).
- Isolated people are discontent, other people are content.
- Some swaps and and their total profit:

Trade	Profit
DD $\rightarrow$ CC	2
DC $\rightarrow$ CC	1
CD $\rightarrow$ DC	0
CC $\rightarrow$ CD	-1
CC $\rightarrow$ DD	-2

Schelling's model in 1D (circle)



- Schelling (1969, 1971, 1978).
- Isolated people are discontent, other people are content.
- Some swaps and and their total profit:

Trade	Profit
DD $\to$ CC	2
DC $\to$ CC	1
CD $\to$ DC	0
CC $\to$ CD	-1
CC $\to$ DD	-2

- This "problem" can be "solved" in "hundreds" of ways. (Analytically, stochastically, whatever.)

Young's take on Schelling's model

Young's take on Schelling's model

- Some trades and their total profit:

Trade	Profit	Probability
DD $\to$ CC	$2 - 2m$	$1 - \mathcal{O}(\epsilon)$
DC $\to$ CC	$1 - 2m$	$1 - \mathcal{O}(\epsilon)$
CD $\to$ DC	$0 - 2m$	low: ϵ^a
CC $\to$ CD	$-1 - 2m$	lower: ϵ^b
CC $\to$ DD	$-2 - 2m$	lowest: ϵ^c

Young's take on Schelling's model

- Some trades and their total profit:

Trade	Profit	Probability
DD $\rightarrow$ CC	$2 - 2m$	$1 - \mathcal{O}(\epsilon)$
DC $\rightarrow$ CC	$1 - 2m$	$1 - \mathcal{O}(\epsilon)$
CD $\rightarrow$ DC	$0 - 2m$	low: ϵ^a
CC $\rightarrow$ CD	$-1 - 2m$	lower: ϵ^b
CC $\rightarrow$ DD	$-2 - 2m$	lowest: ϵ^c

Where m are moving costs and $0 < a < b < c$.

Young's take on Schelling's model

- Some trades and their total profit:

Trade	Profit	Probability
DD $\rightarrow$ CC	$2 - 2m$	$1 - \mathcal{O}(\epsilon)$
DC $\rightarrow$ CC	$1 - 2m$	$1 - \mathcal{O}(\epsilon)$
CD $\rightarrow$ DC	$0 - 2m$	low: ϵ^a
CC $\rightarrow$ CD	$-1 - 2m$	lower: ϵ^b
CC $\rightarrow$ DD	$-2 - 2m$	lowest: ϵ^c

  Where m are moving costs and $0 < a < b < c$.

- P^0 only allows beneficial trades.

Young's take on Schelling's model

- Some trades and their total profit:

Trade	Profit	Probability
DD $\to$ CC	$2 - 2m$	$1 - \mathcal{O}(\epsilon)$
DC $\to$ CC	$1 - 2m$	$1 - \mathcal{O}(\epsilon)$
CD $\to$ DC	$0 - 2m$	low: ϵ^a
CC $\to$ CD	$-1 - 2m$	lower: ϵ^b
CC $\to$ DD	$-2 - 2m$	lowest: ϵ^c

Where m are moving costs and $0 < a < b < c$.

- P^0 only allows beneficial trades.
- P^ϵ allows all trades.

Young's take on Schelling's model

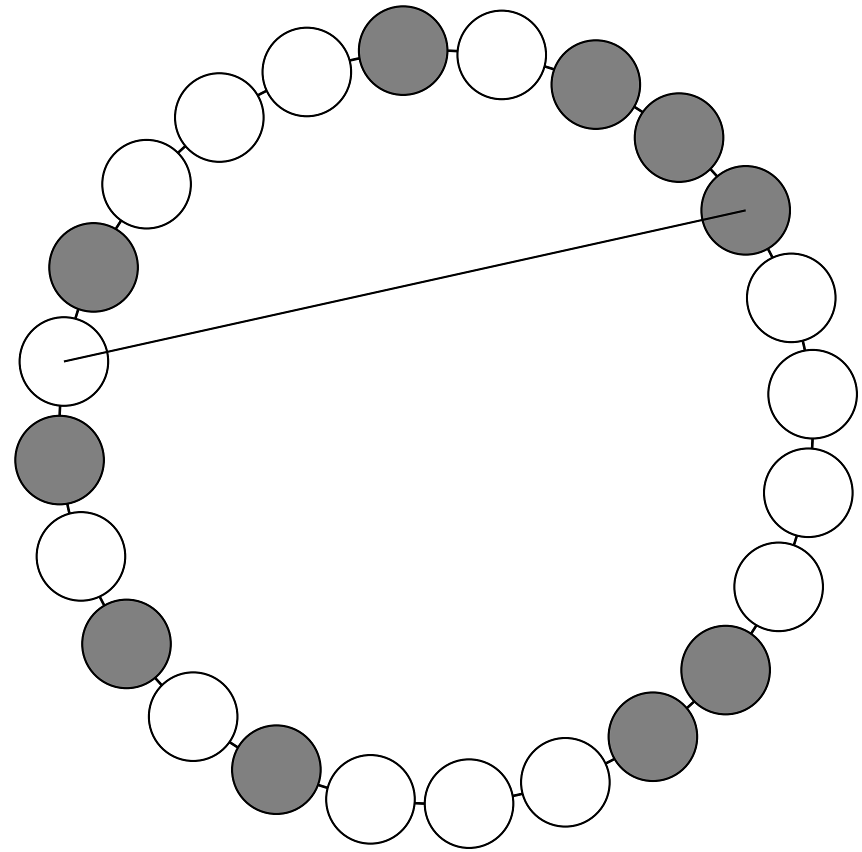
- Some trades and their total profit:

Trade	Profit	Probability
DD $\to$ CC	$2-2m$	$1-\mathcal{O}(\epsilon)$
DC $\to$ CC	$1-2m$	$1-\mathcal{O}(\epsilon)$
CD $\to$ DC	$0-2m$	low: ϵ^a
CC $\to$ CD	$-1-2m$	lower: ϵ^b
CC $\to$ DD	$-2-2m$	lowest: ϵ^c

Where m are moving costs and $0 < a < b < c$.

- P^0 only allows beneficial trades.

- P^ϵ allows all trades. P^ϵ is ergodic.



An example of a trade with cost $1-2m$ and probability 1 (in P^0) or near 1 (in P^ϵ).

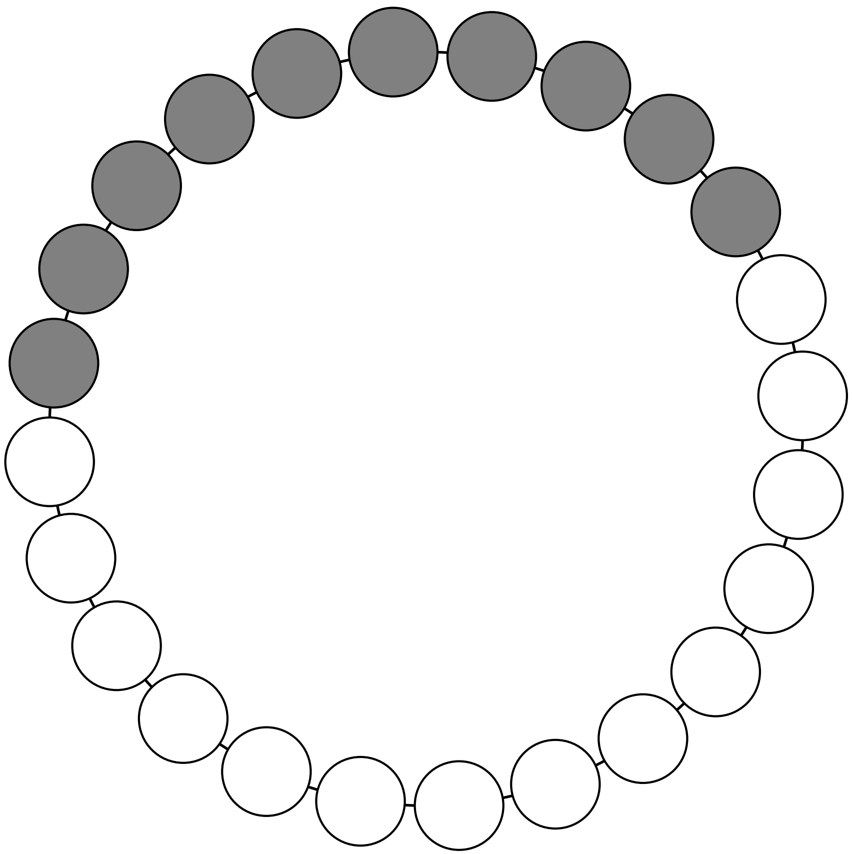
Stochastically stable = completely segregated

Stochastically stable = completely segregated

- Absorbing states in P^0 are either completely segregated or dispersed: $A = S \cup D$.

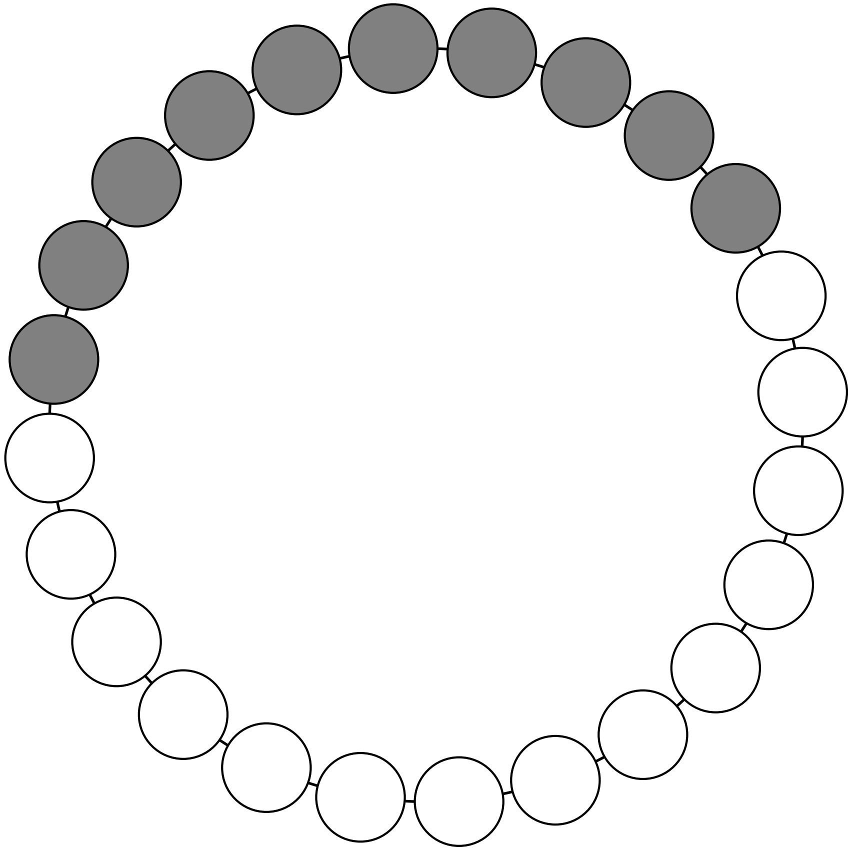
Stochastically stable = completely segregated

- Absorbing states in P^0 are either completely segregated or dispersed: $A = S \cup D$.



Stochastically stable = completely segregated

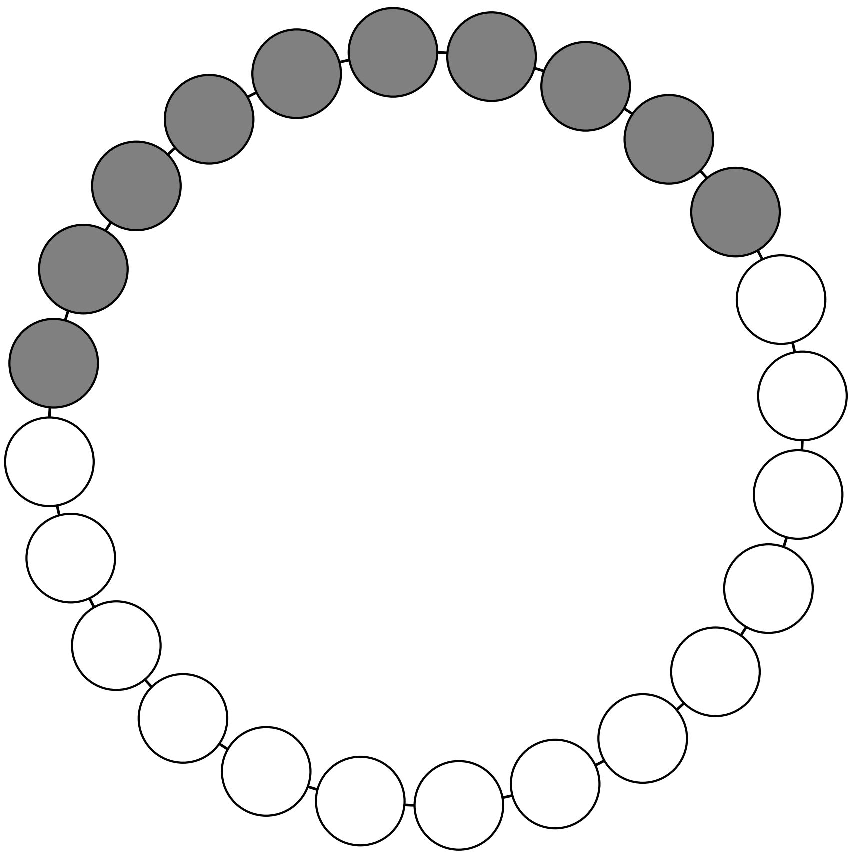
- Absorbing states in P^0 are either completely segregated or dispersed: $A = S \cup D$.



- Completely segregated $=_{Def}$ a state in which all the A's are lined up on one side of the circle and all the B's are on the other.

Stochastically stable = completely segregated

- Absorbing states in P^0 are either completely segregated or dispersed: $A = S \cup D$.



- Completely segregated $=_{Def}$ a state in which all the A's are lined up on one side of the circle and all the B's are on the other.
- Dispersed $=_{Def}$ absorbing nut not completely segregated.

Theorem (H. Peyton Young, 1993). The set of stochastically stable states in the perturbed one-dimensional version of Schelling's model equals the set of completely segregated absorbing states S.

Determine all Recurrence classes of P_0

Determine all Recurrence classes of P_0

- All absorbing states A are recurrent.

Determine all Recurrence classes of P_0

- All absorbing states A are recurrent.
- If not absorbing, then a mutually advantageous swap with another state possible.

Determine all Recurrence classes of P_0

- All absorbing states A are recurrent.
- If not absorbing, then a mutually advantageous swap with another state possible. (Exchange places with an associate's predecessor.)

Determine all Recurrence classes of P_0

- All absorbing states A are recurrent.
- If not absorbing, then a mutually advantageous swap with another state possible. (Exchange places with an associate's predecessor.)
  - The total discontent decreases.

Determine all Recurrence classes of P_0

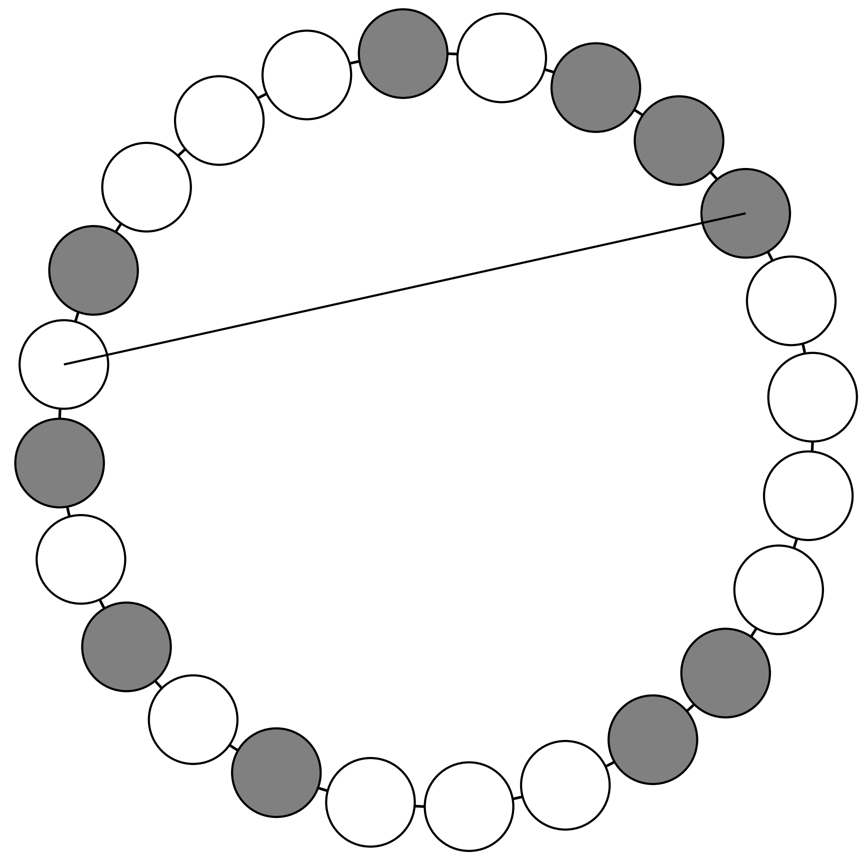
- All absorbing states A are recurrent.
- If not absorbing, then a mutually advantageous swap with another state possible. (Exchange places with an associate's predecessor.)
  - The total discontent decreases.
  - By repeating this step it is possible to travel from every state to an absorbing state.

Determine all Recurrence classes of P_0

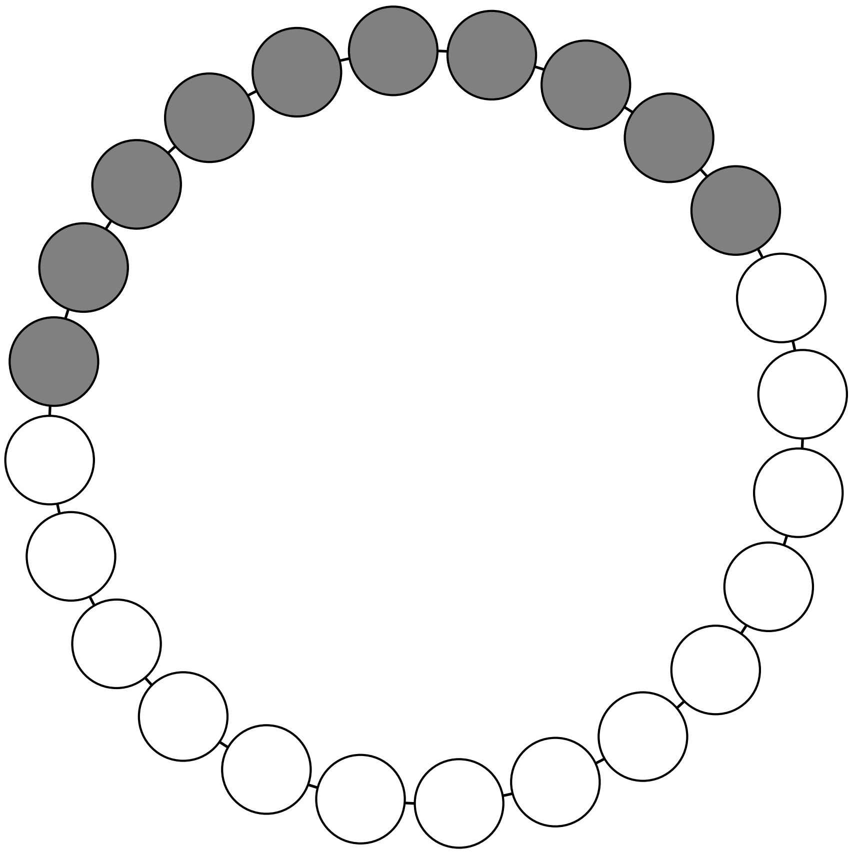
- All absorbing states A are recurrent.
- If not absorbing, then a mutually advantageous swap with another state possible. (Exchange places with an associate's predecessor.)
  - The total discontent decreases.
  - By repeating this step it is possible to travel from every state to an absorbing state.
- So all and only Recurrence classes are singletons of absorbing states: $\mathcal{R} = \{\{a\} \mid a \in A\}$.

Determine all Recurrence classes of P_0

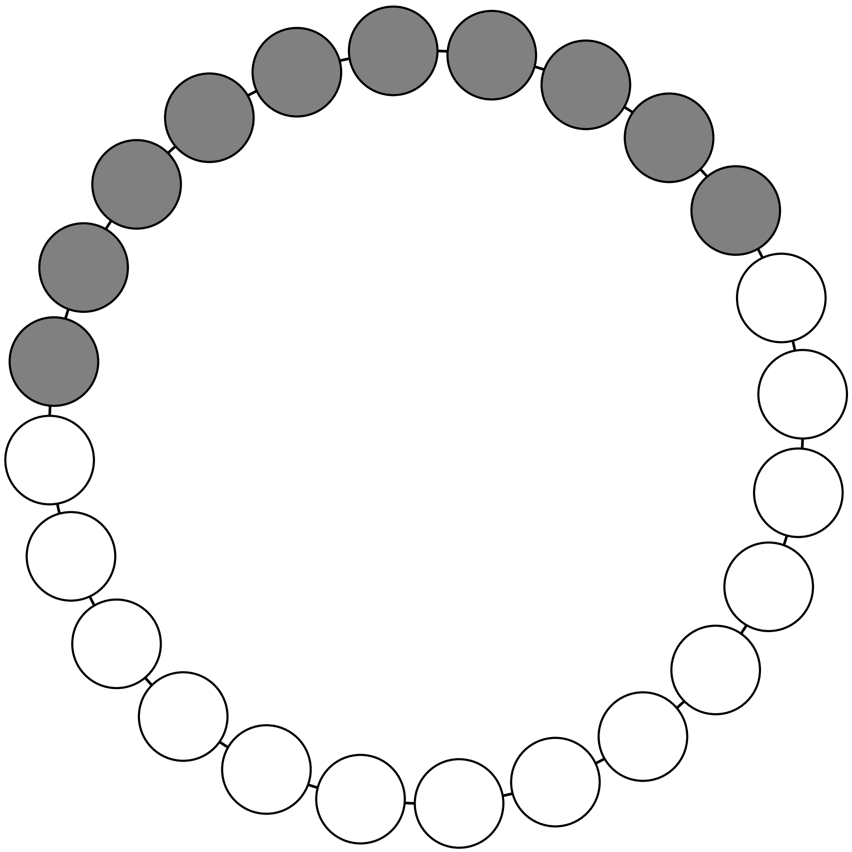
- All absorbing states A are recurrent.
- If not absorbing, then a mutually advantageous swap with another state possible. (Exchange places with an associate's predecessor.)
  - The total discontent decreases.
  - By repeating this step it is possible to travel from every state to an absorbing state.
- So all and only Recurrence classes are singletons of absorbing states: $\mathcal{R} = \{\{a\} \mid a \in A\}$.



- Absorbing states are either completely segregated or dispersed: $A = S \cup D$.

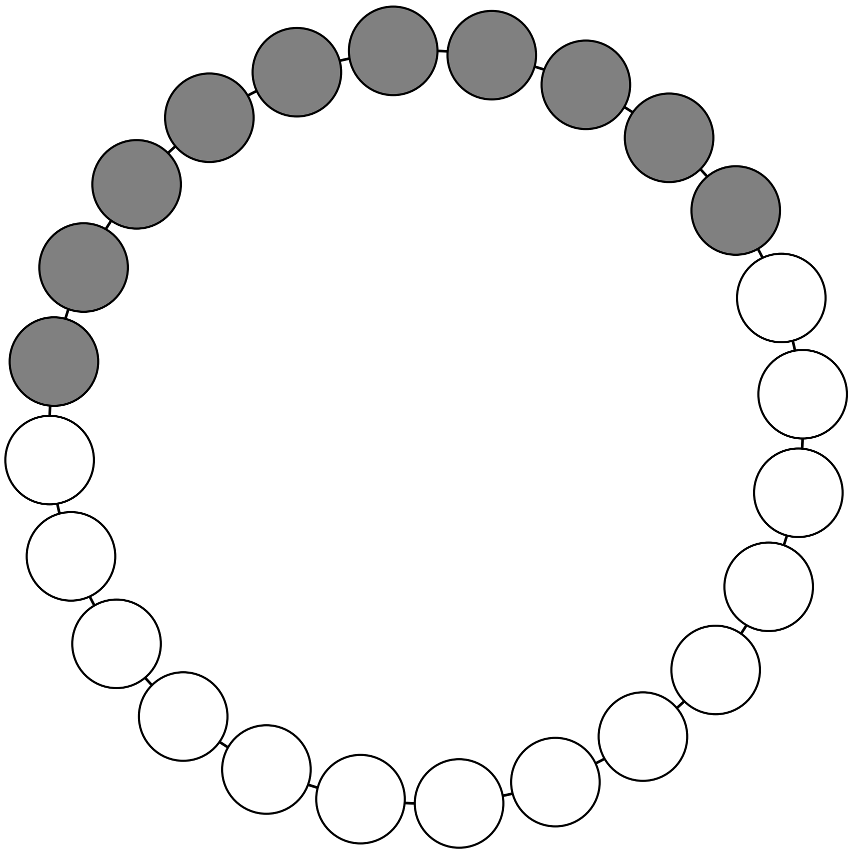


- Absorbing states are either completely segregated or dispersed: $A = S \cup D$.



- For all $z, z' \in A$, let $r(z \to z')$ be defined as usual (the least resistance from z to z').

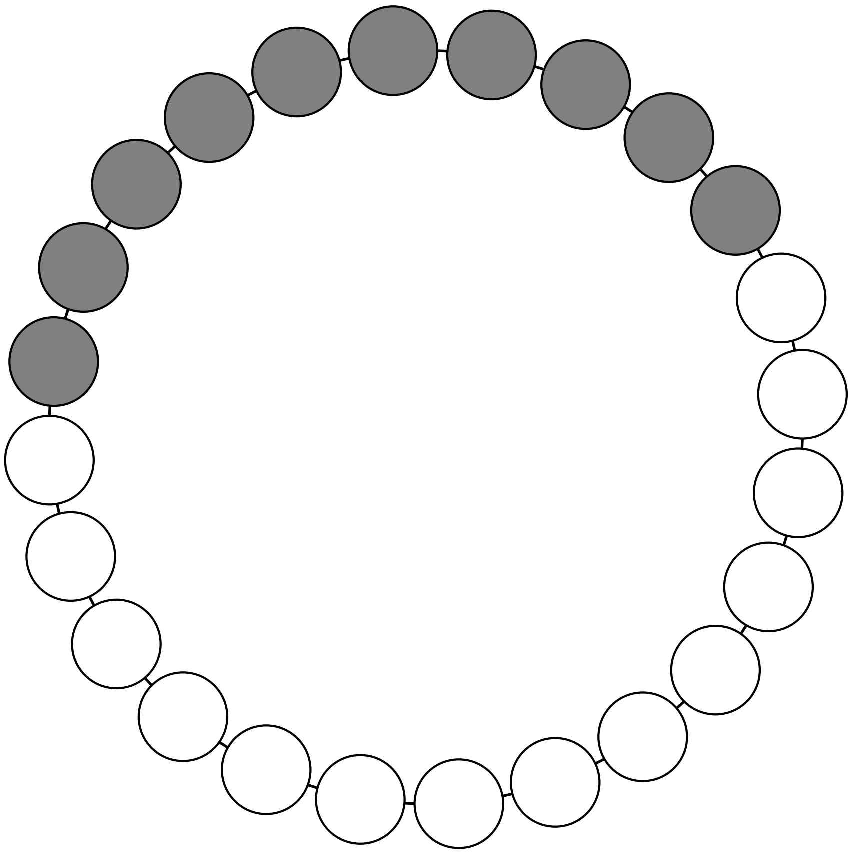
- Absorbing states are either completely segregated or dispersed: $A = S \cup D$.



- For all $z, z' \in A$, let $r(z \to z')$ be defined as usual (the least resistance from z to z').

Claims:

- Absorbing states are either completely segregated or dispersed: $A = S \cup D$.

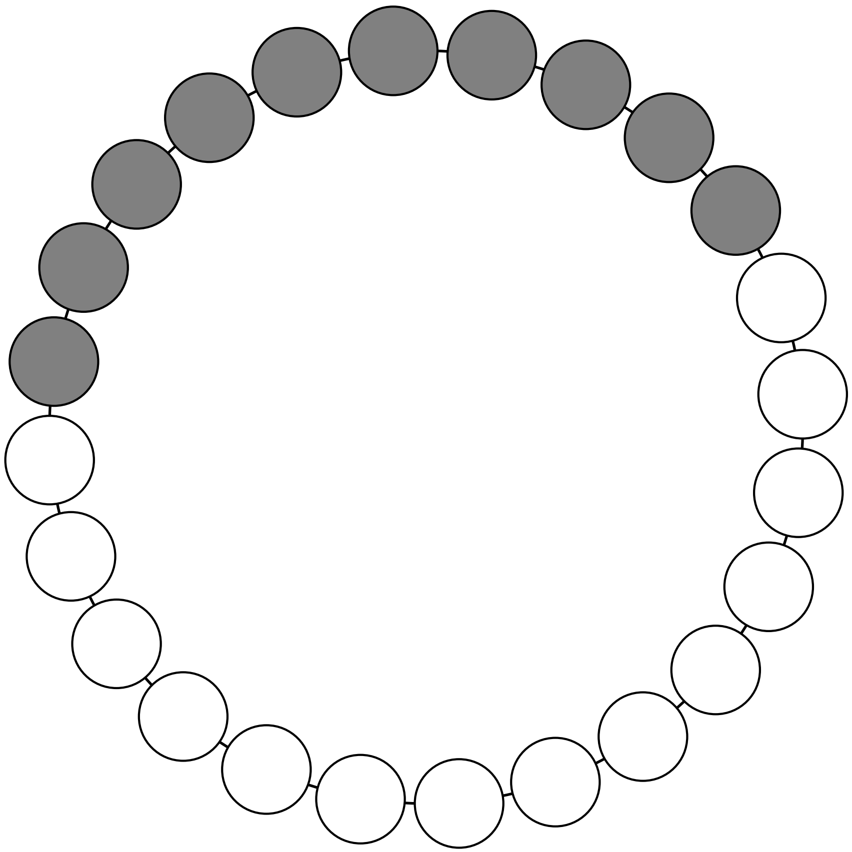


- For all $z, z' \in A$, let $r(z \to z')$ be defined as usual (the least resistance from z to z').

Claims:

1. If $s \in D$, there **does not** exist an s-tree from $A \backslash \{s\}$ with only a-edges.

- Absorbing states are either completely segregated or dispersed: $A = S \cup D$.

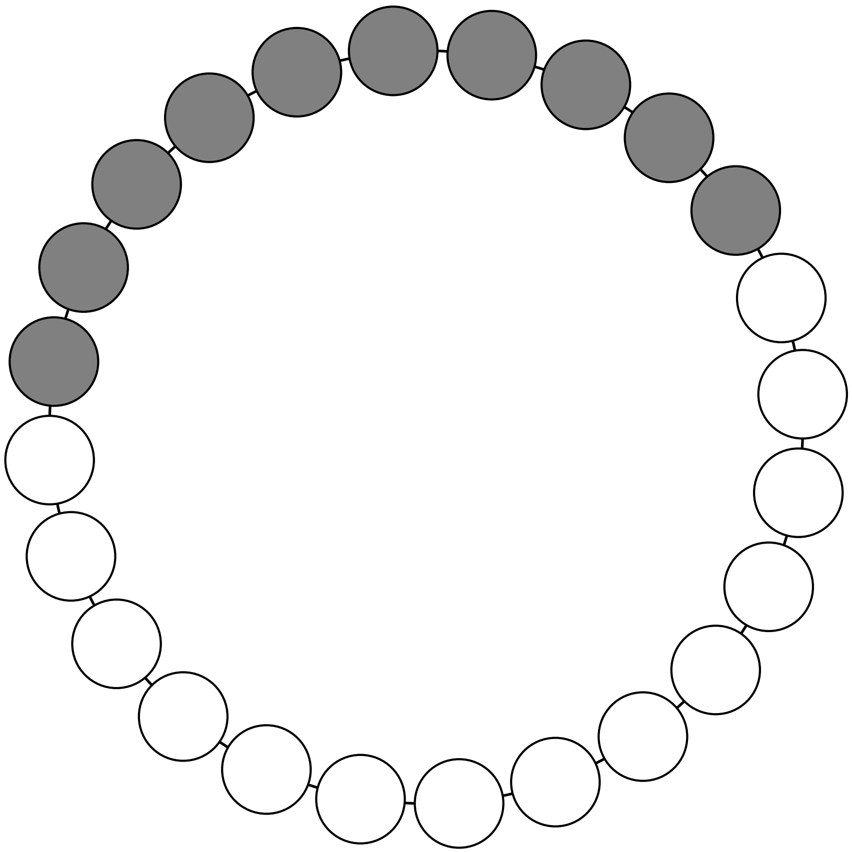


- For all $z, z' \in A$, let $r(z \to z')$ be defined as usual (the least resistance from z to z').

Claims:

1. If $s \in D$, there **does not** exist an s-tree from $A \backslash \{s\}$ with only a-edges.
2. If $s \in S$, there **does** exist an s-tree from $A \backslash \{s\}$ with only a-edges.

- Absorbing states are either completely segregated or dispersed: $A = S \cup D$.



- For all $z, z' \in A$, let $r(z \to z')$ be defined as usual (the least resistance from z to z').

Claims:

1. If $s \in D$, there **does not** exist an s-tree from $A \backslash \{s\}$ with only a-edges.

2. If $s \in S$, there **does** exist an s-tree from $A \backslash \{s\}$ with only a-edges.

3. The classes with lowest potential are

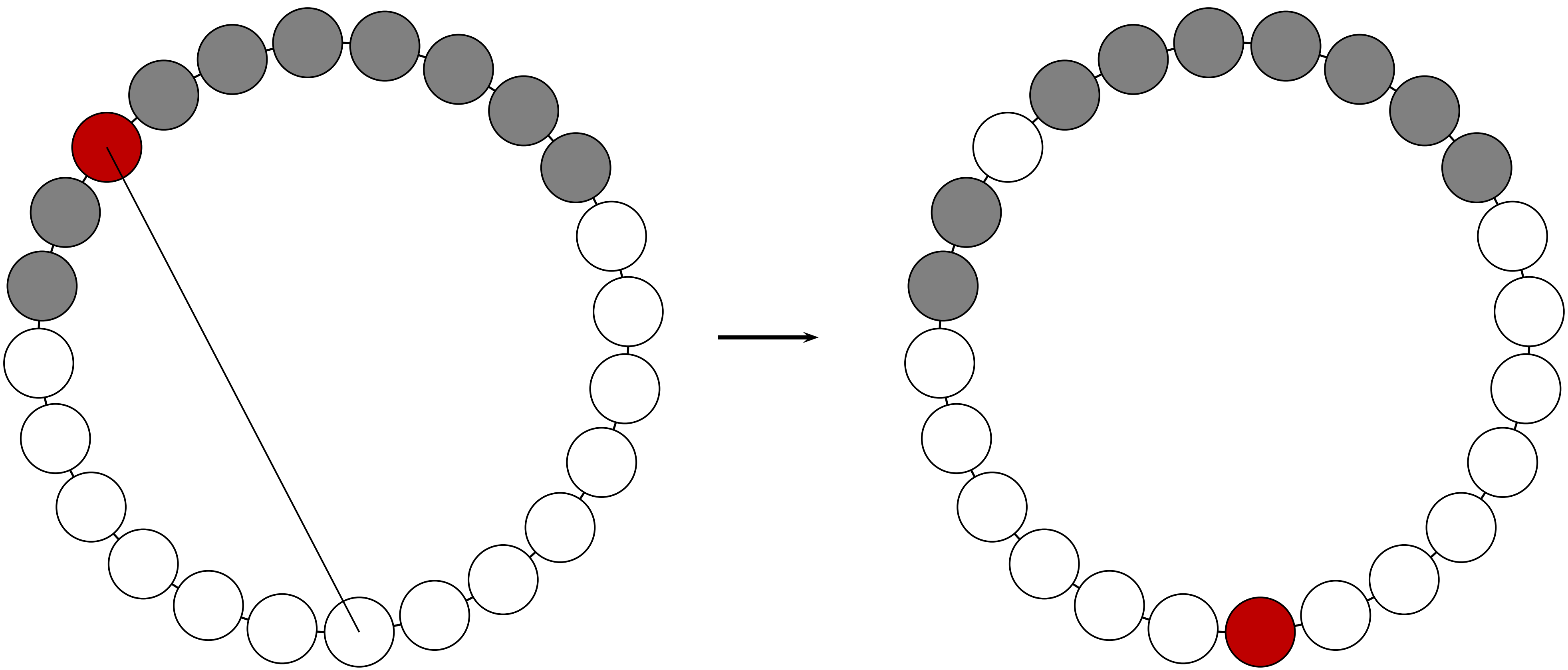
$$\mathcal{L} = \{\{s\} \mid s \in S\}.$$

Claim 1

If $s \in D$, there **does not** exist an s-tree from $A\backslash\{s\}$ with only a-edges.

Claim 1

If $s \in D$, there **does not** exist an s-tree from $A \backslash \{s\}$ with only a-edges.

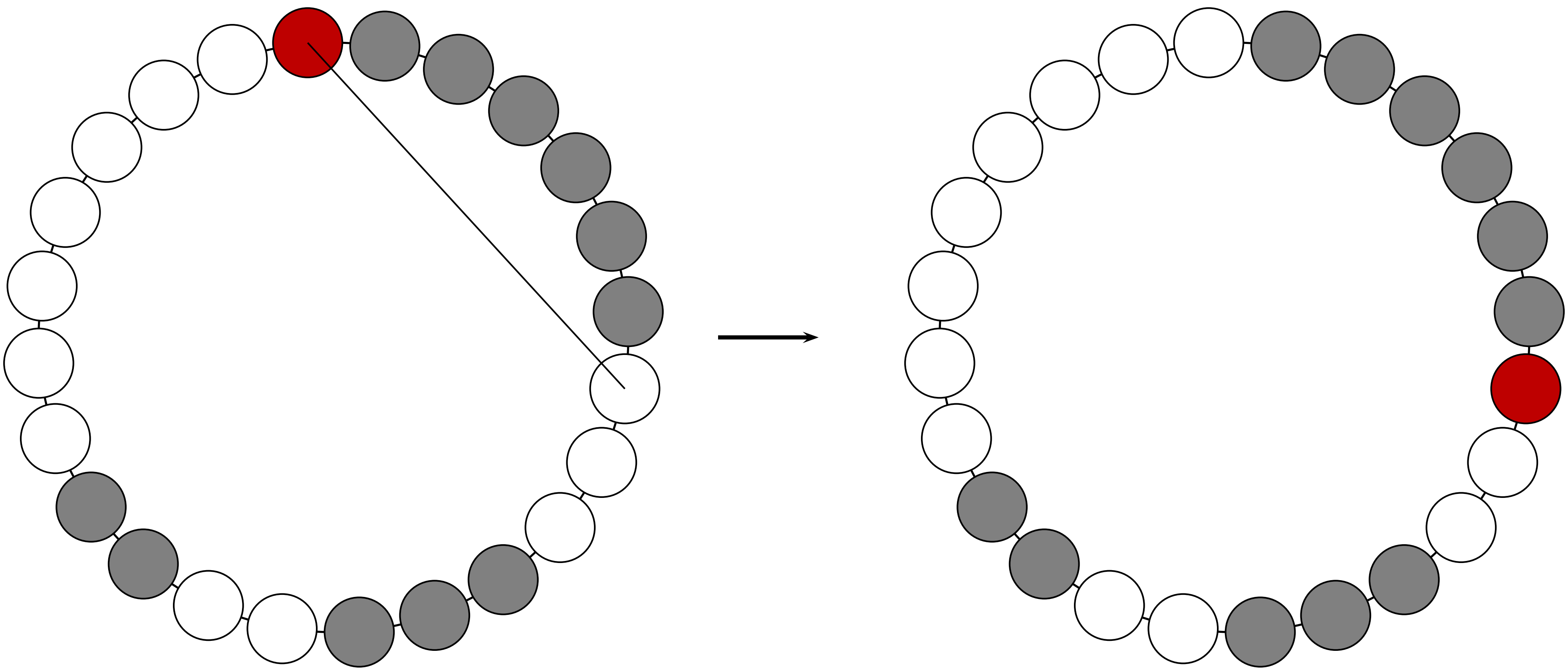


Claim 2: a resistance

If $s \in S$, there **does** exist an s-tree from $A \backslash \{s\}$ with only a-edges.

Claim 2: a resistance

If $s \in S$, there **does** exist an s-tree from $A \backslash \{s\}$ with only a-edges.

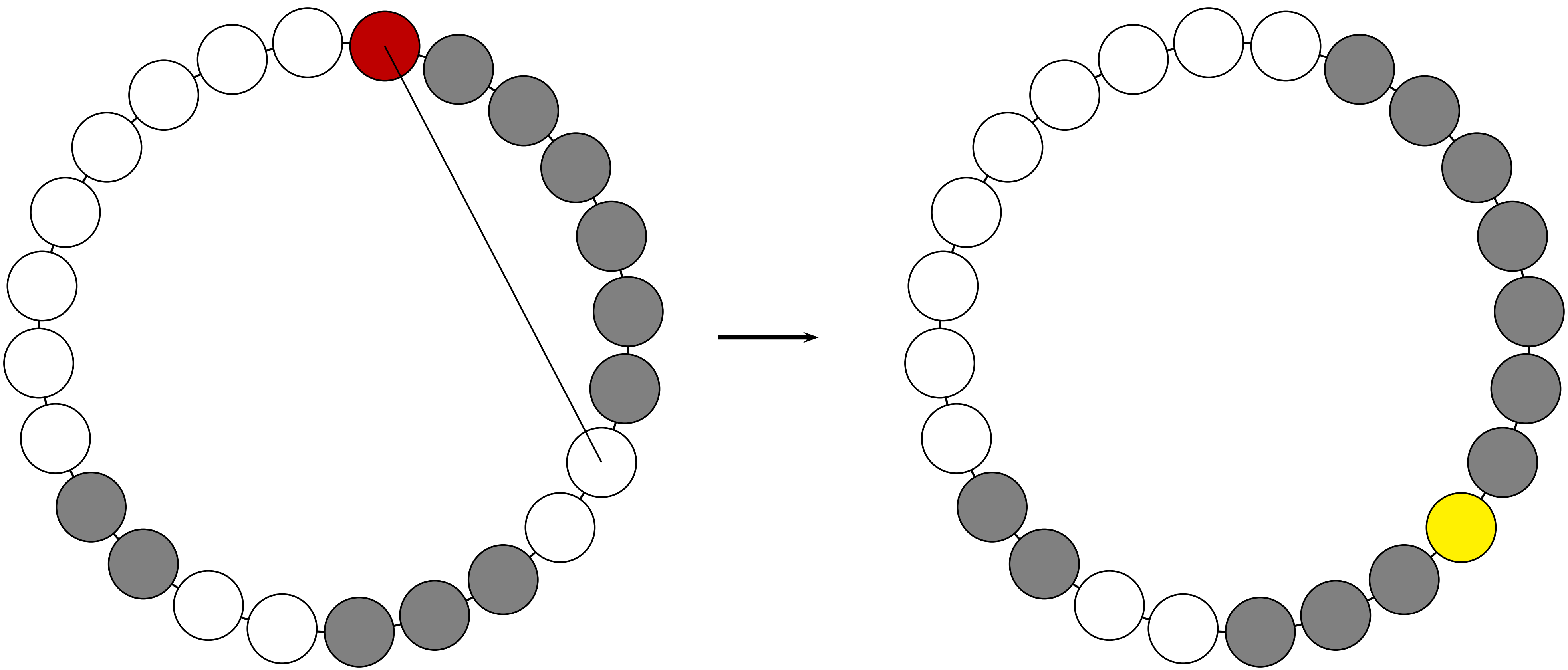


Claim 2: a resistance, discontent individual

If $s \in S$, there **does** exist an s-tree from $A \backslash \{s\}$ with only a-edges.

Claim 2: a resistance, discontent individual

If $s \in S$, there **does** exist an s-tree from $A \backslash \{s\}$ with only a-edges.



Absorbing state $\Leftrightarrow$ state with low potential

Claim 1: *If $s \in D$, there **does not** exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

Absorbing state $\Leftrightarrow$ state with low potential

Claim 1: *If $s \in D$, there **does not** exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

- Let $s \in D$. We must show that **some** edges from A to s have resistance $> a$.

Absorbing state $\Leftrightarrow$ state with low potential

Claim 1: *If $s \in D$, there **does not** exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

- Let $s \in D$. We must show that **some** edges from A to s have resistance $> a$.
- Edges from S to s, at least, necessarily involve moves that create at least one discontent (= isolated) individual.

Absorbing state $\Leftrightarrow$ state with low potential

Claim 1: *If $s \in D$, there **does not** exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

- Let $s \in D$. We must show that **some** edges from A to s have resistance $> a$.
- Edges from S to s, at least, necessarily involve moves that create at least one discontent (= isolated) individual.
- Therefore, all s-trees in A have resistance strictly larger than $a(|A| - 1)$.

Absorbing state $\Leftrightarrow$ state with low potential

Claim 1: *If $s \in D$, there **does not** exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

- Let $s \in D$. We must show that **some** edges from A to s have resistance $> a$.
- Edges from S to s, at least, necessarily involve moves that create at least one discontent (= isolated) individual.
- Therefore, all s-trees in A have resistance strictly larger than $a(|A| - 1)$.

Claim 2: *If $s \in S$, there **does** exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

Absorbing state $\Leftrightarrow$ state with low potential

Claim 1: *If $s \in D$, there* ***does not*** *exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

- Let $s \in D$. We must show that **some** edges from A to s have resistance $> a$.
- Edges from S to s, at least, necessarily involve moves that create at least one discontent (= isolated) individual.
- Therefore, all s-trees in A have resistance strictly larger than $a(|A| - 1)$.

Claim 2: *If $s \in S$, there* ***does*** *exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

- Let $s \in S$. We must show that **all** edges from A to s have resistance a.

Absorbing state ⇔ state with low potential

Claim 1: *If $s \in D$, there **does not** exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

- Let $s \in D$. We must show that **some** edges from A to s have resistance $> a$.
- Edges from S to s, at least, necessarily involve moves that create at least one discontent (= isolated) individual.
- Therefore, all s-trees in A have resistance strictly larger than $a(|A| - 1)$.

Claim 2: *If $s \in S$, there **does** exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

- Let $s \in S$. We must show that **all** edges from A to s have resistance a.
- From elements in S to other elements in S: ok! (Put tail to head repeatedly.)

Absorbing state $\Leftrightarrow$ state with low potential

Claim 1: *If $s \in D$, there **does not** exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

- Let $s \in D$. We must show that **some** edges from A to s have resistance $> a$.
- Edges from S to s, at least, necessarily involve moves that create at least one discontent (= isolated) individual.
- Therefore, all s-trees in A have resistance strictly larger than $a(|A| - 1)$.

Claim 2: *If $s \in S$, there **does** exist an s-tree from $A \backslash \{s\}$ with only a-edges.*

- Let $s \in S$. We must show that **all** edges from A to s have resistance a.
- From elements in S to other elements in S: ok! (Put tail to head repeatedly.)
- From elements in D to elements in S: ok! (Put tail to head of small groups repeatedly. If one large cluster, continue as in previous case.)

Summary of Young's take on Schelling's model

Theorem (H. Peyton Young, 1993). The set of stochastically stable states in the perturbed one-dimensional version of Schelling's model equals the set of completely segregated absorbing states S.

Summary of Young's take on Schelling's model

Theorem (H. Peyton Young, 1993). The set of stochastically stable states in the perturbed one-dimensional version of Schelling's model equals the set of completely segregated absorbing states S.

- If minimal random behaviour is allowed, "sub-optimal" stable states ascend into "global optimal" stable states.

Summary of Young's take on Schelling's model

Theorem (H. Peyton Young, 1993). The set of stochastically stable states in the perturbed one-dimensional version of Schelling's model equals the set of completely segregated absorbing states S.

- If minimal random behaviour is allowed, "sub-optimal" stable states ascend into "global optimal" stable states. In the case of (Young's take on) Schelling's model:

Summary of Young's take on Schelling's model

Theorem (H. Peyton Young, 1993). The set of stochastically stable states in the perturbed one-dimensional version of Schelling's model equals the set of completely segregated absorbing states S.

- If minimal random behaviour is allowed, "sub-optimal" stable states ascend into "global optimal" stable states. In the case of (Young's take on) Schelling's model:
  - If agents act **irrational most of the time**, no segregation will occur.

Summary of Young's take on Schelling's model

Theorem (H. Peyton Young, 1993). The set of stochastically stable states in the perturbed one-dimensional version of Schelling's model equals the set of completely segregated absorbing states S.

- If minimal random behaviour is allowed, "sub-optimal" stable states ascend into "global optimal" stable states. In the case of (Young's take on) Schelling's model:
  - If agents act **irrational most of the time**, no segregation will occur.
  - If agents act **rational all of the time**, modest forms of segregation will occur (P^0).

Summary of Young's take on Schelling's model

Theorem (H. Peyton Young, 1993). The set of stochastically stable states in the perturbed one-dimensional version of Schelling's model equals the set of completely segregated absorbing states S.

- If minimal random behaviour is allowed, "sub-optimal" stable states ascend into "global optimal" stable states. In the case of (Young's take on) Schelling's model:
  - If agents act **irrational most of the time**, no segregation will occur.
  - If agents act **rational all of the time**, modest forms of segregation will occur (P^0).
  - If agents act **rational most of the time, but not always**, more severe forms of segregation will occur (P^ϵ).

Summary of Young's take on Schelling's model

Theorem (H. Peyton Young, 1993). The set of stochastically stable states in the perturbed one-dimensional version of Schelling's model equals the set of completely segregated absorbing states S.

- If minimal random behaviour is allowed, "sub-optimal" stable states ascend into "global optimal" stable states. In the case of (Young's take on) Schelling's model:
  - If agents act **irrational most of the time**, no segregation will occur.
  - If agents act **rational all of the time**, modest forms of segregation will occur (P^0).
  - If agents act **rational most of the time, but not always**, more severe forms of segregation will occur (P^ϵ).
- Conclusion: if you want to have a nice and varied society do not think too much about where you want to live

Summary of Young's take on Schelling's model

Theorem (H. Peyton Young, 1993). The set of stochastically stable states in the perturbed one-dimensional version of Schelling's model equals the set of completely segregated absorbing states S.

- If minimal random behaviour is allowed, "sub-optimal" stable states ascend into "global optimal" stable states. In the case of (Young's take on) Schelling's model:
  - If agents act **irrational most of the time**, no segregation will occur.
  - If agents act **rational all of the time**, modest forms of segregation will occur (P^0).
  - If agents act **rational most of the time, but not always**, more severe forms of segregation will occur (P^ϵ).
- Conclusion: if you want to have a nice and varied society do not think too much about where you want to live : -)