Time-Varying Transition Probability Matrix

Application to Brand Share Analysis

Noboru Murata

June 20, 2023

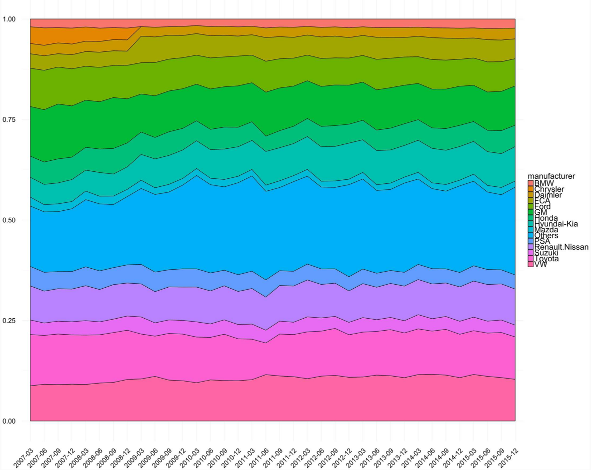
https://noboru-murata.github.io/

Introduction

two problems in graph inference:

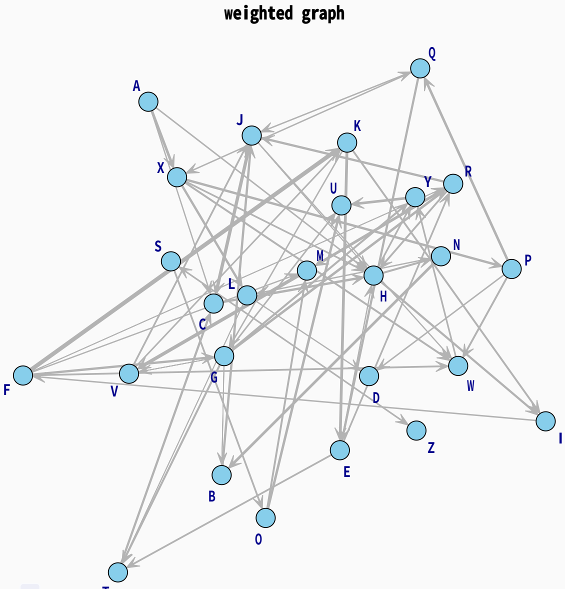
- direct problem (deductive):
  induce properties from known graph structure
  - shortest path, traveling salesman, graph coloring %problem
  - maximum flow, minimum cut %problem
  - Google search (stationary distribution of Google matrix)
- inverse problem (inductive):
  estimate graph structure from partially observed properties
  - path analysis, graphical model estimation
  - sparse estimation of accuracy matrix
  - structure estimation via graph Laplacian

automobile sales data of manufacturers



Questions

- why sales shares vary?
- what happens in customer preferences?



directed graph

- number of node: 26
- edge exist.: 0.1
- weight: uniform on $[0, 1]$

- adjacency matrix W

$$(W)_{ij} = \text{strength of connection from } i \text{ to } j$$

- indicate vector of sink node $\boldsymbol{a}$

$$(\boldsymbol{a})_i = \begin{cases} 1, & \text{if } (W\boldsymbol{e})_i = 0 \\ 0, & \text{otherwise} \end{cases}$$

where $\boldsymbol{e}$ is a vector of all 1

- normalized adjacency matrix (transition matrix) H

$$H = \mathrm{diag}(W\boldsymbol{e} + \boldsymbol{a})^{-1} W$$

$(H\boldsymbol{e})_i = 1$ holds except for sink nodes

- transition matrix with sink node adjustment S

$$\begin{aligned} S =& H + \boldsymbol{a}\boldsymbol{e}^T/n \\ =& (\text{probabilistic transition}) \\ &+ (\text{escape from sink nodes}) \end{aligned}$$

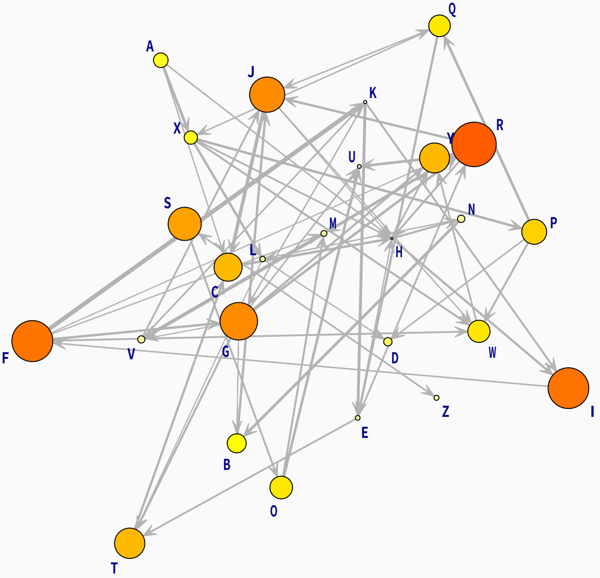
probability matrix: $S\boldsymbol{e} = \boldsymbol{e}$

- Google matrix G

$$\begin{aligned} G =& \alpha S + (1-\alpha)\boldsymbol{e}\boldsymbol{e}^T/n \\ =& (\text{transition along edges}) + (\text{random transition}) \end{aligned}$$

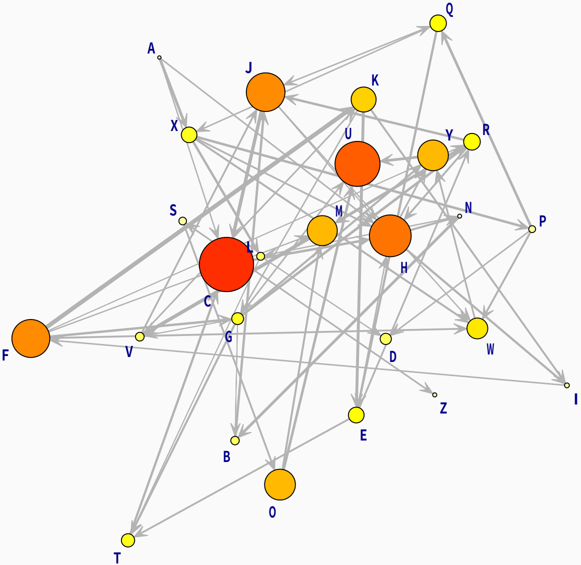
probability matrix: $G\boldsymbol{e} = \boldsymbol{e}$

t = 1



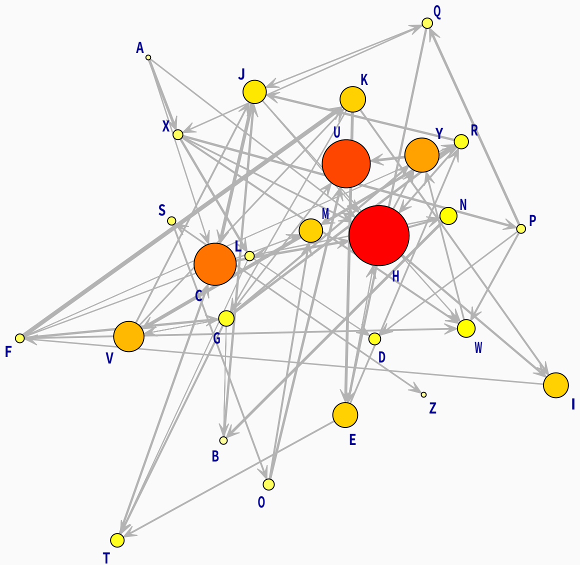
- $\alpha = 0.85$
- initial: \ uniform dist.

t = 2

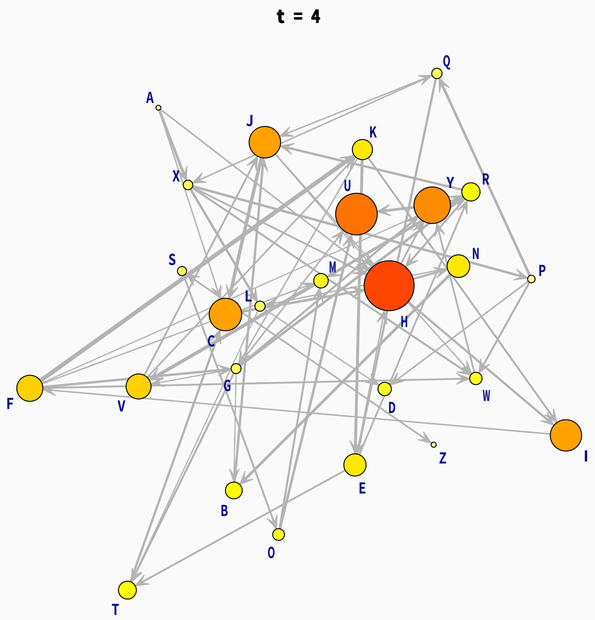


- $\alpha = 0.85$
- initial: \ uniform dist.

t = 3

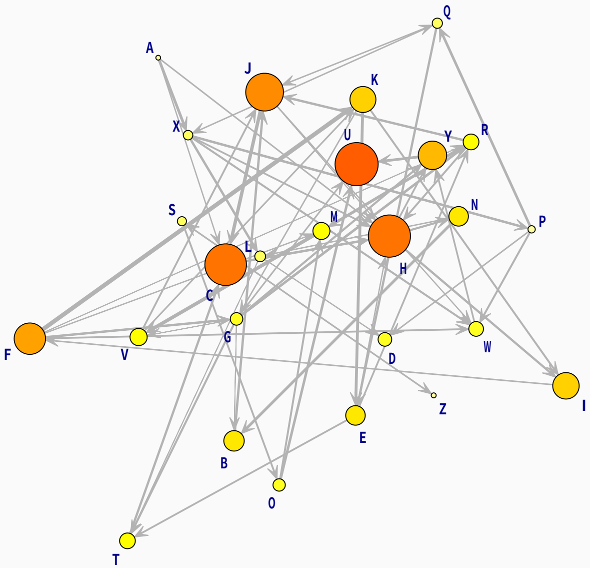


- $\alpha = 0.85$
- initial: \ uniform dist.



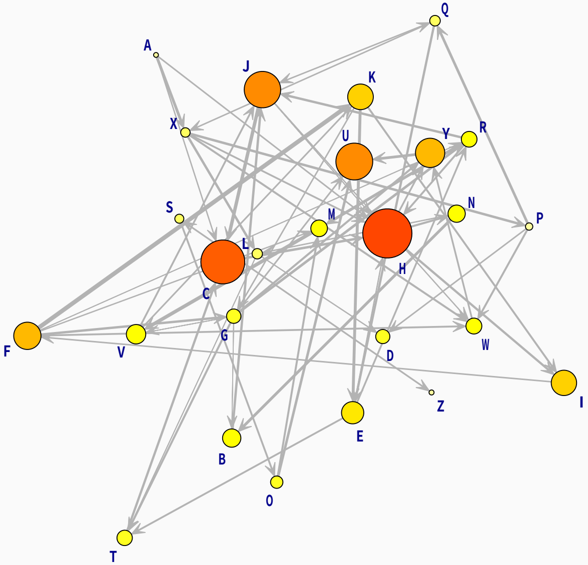
- $\alpha = 0.85$
- initial: \ uniform dist.

t = 5

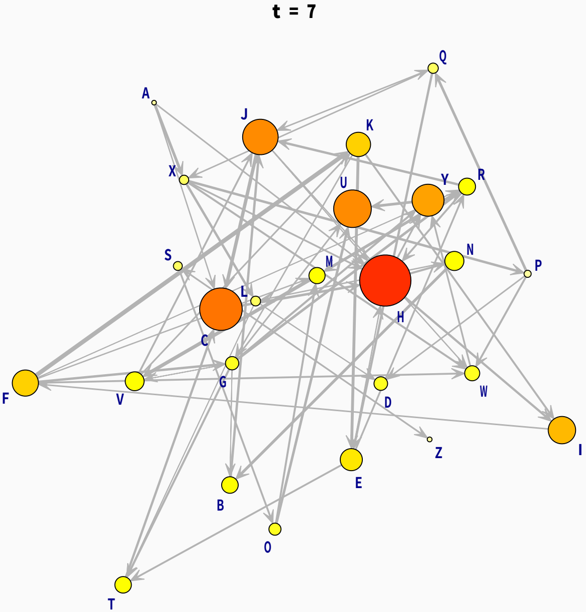


- $\alpha = 0.85$
- initial: \ uniform dist.

t = 6

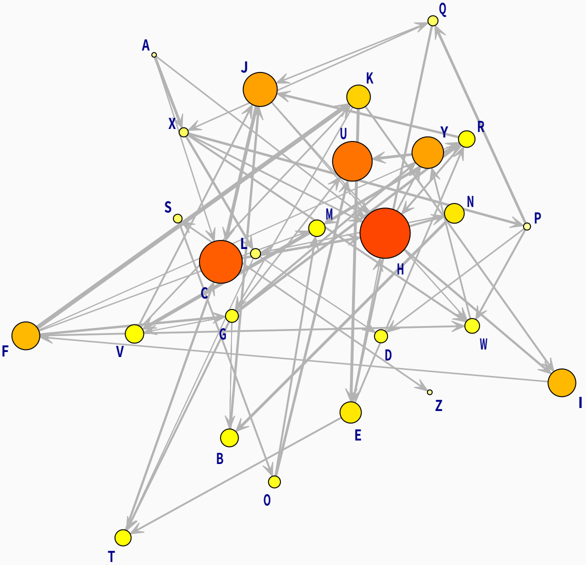


- $\alpha = 0.85$
- initial: \ uniform dist.

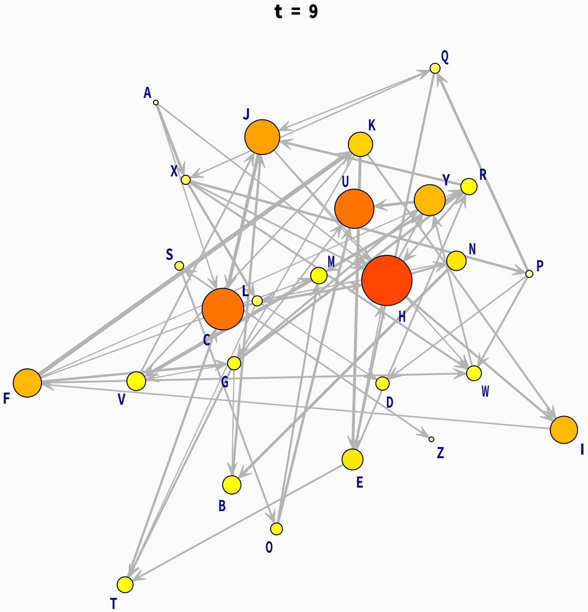


- $\alpha = 0.85$
- initial: \ uniform dist.

t = 8

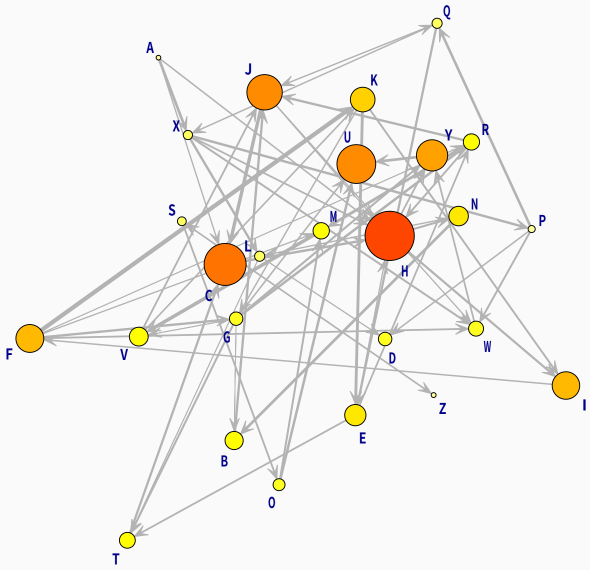


- $\alpha = 0.85$
- initial: \ uniform dist.

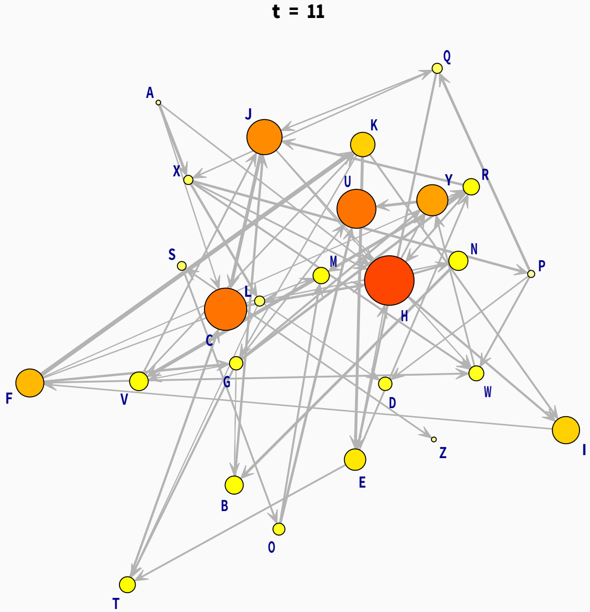


- $\alpha = 0.85$
- initial: \ uniform dist.

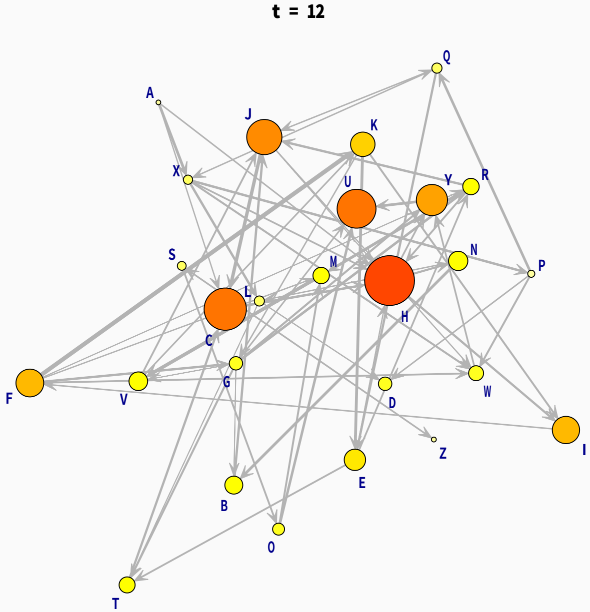
t = 10



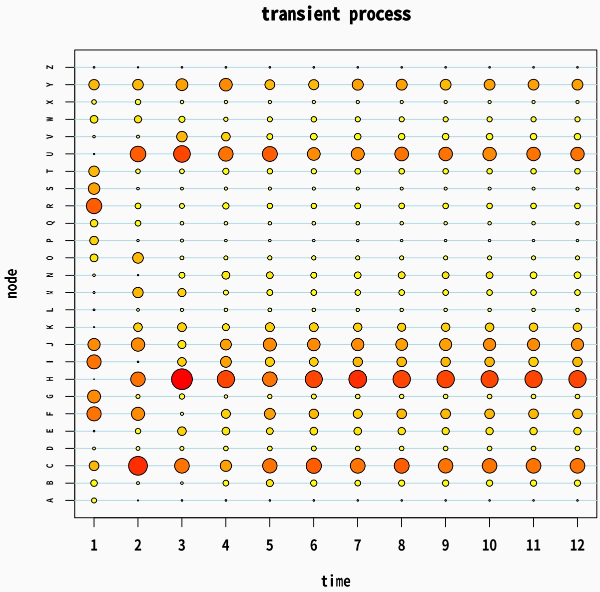
- $\alpha = 0.85$
- initial: \ uniform dist.



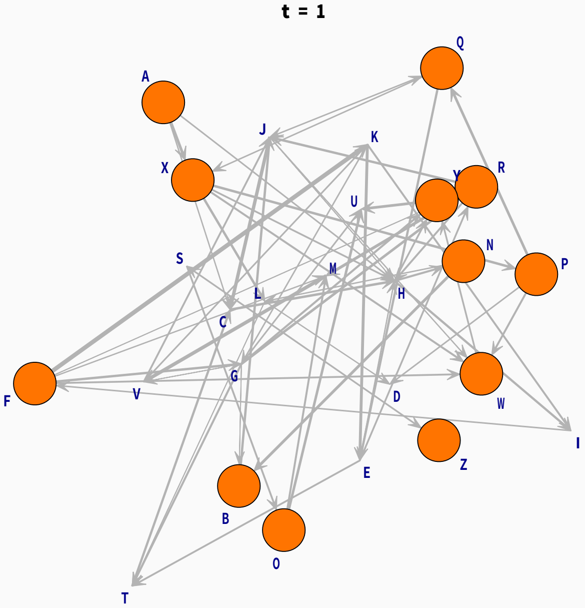
- $\alpha = 0.85$
- initial: \ uniform dist.



- $\alpha = 0.85$
- initial: \ uniform dist.

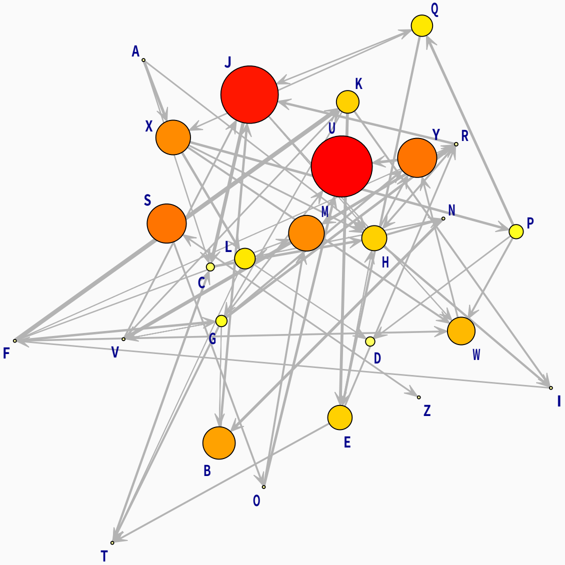


- $\alpha = 0.85$
- initial: \ uniform dist.



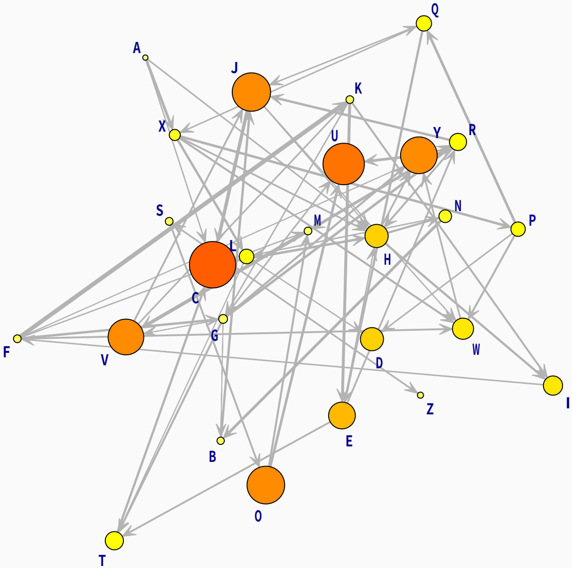
- $\alpha = 0.85$
- initial: 12 nodes

t = 2



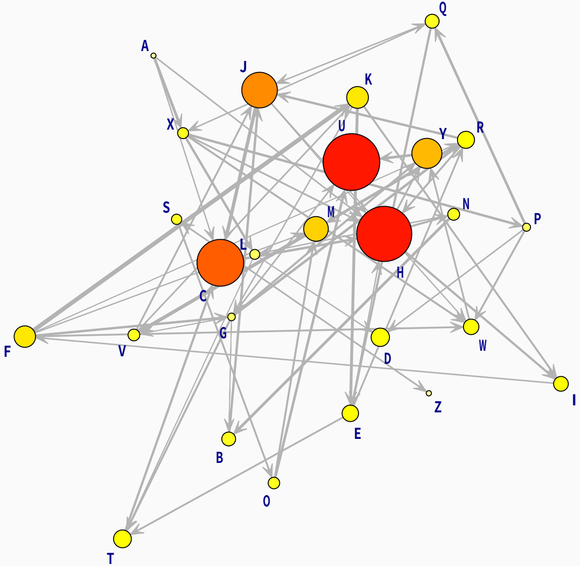
- $\alpha = 0.85$
- initial: 12 nodes

t = 3



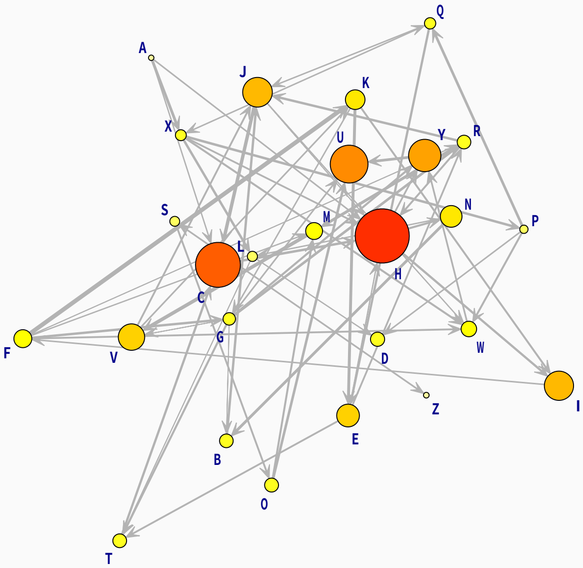
- $\alpha = 0.85$
- initial: 12 nodes

t = 4

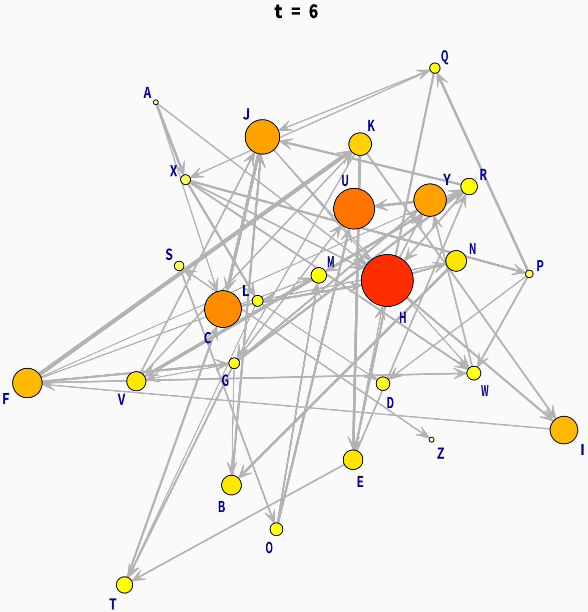


- $\alpha = 0.85$
- initial:
  12 nodes

t = 5

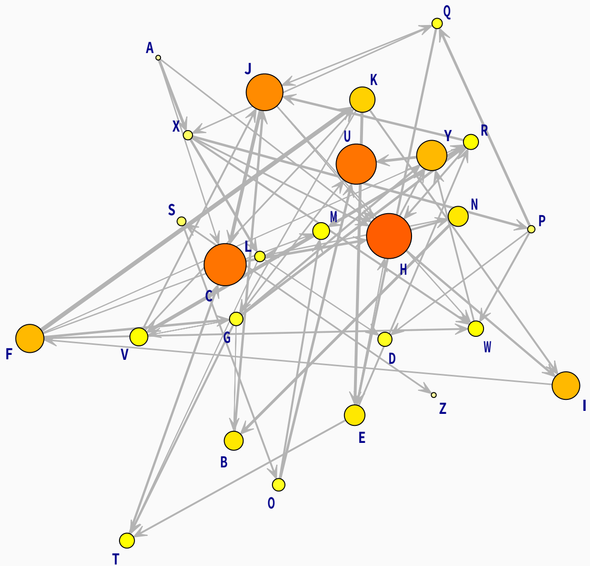


- $\alpha = 0.85$
- initial: 12 nodes

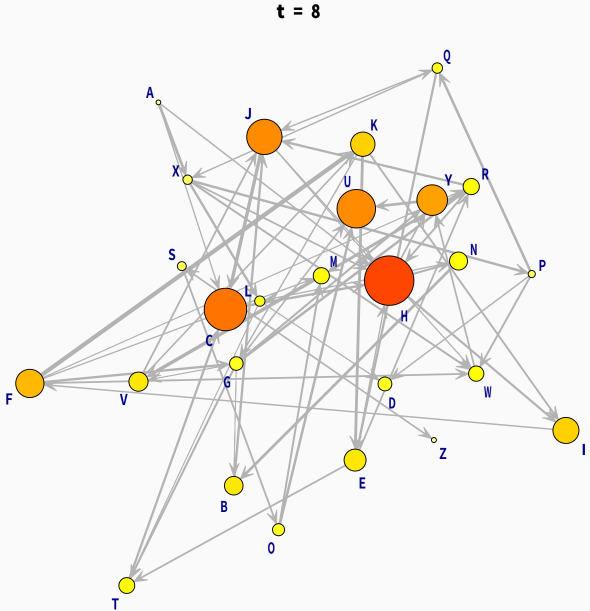


- $\alpha = 0.85$
- initial:
  12 nodes

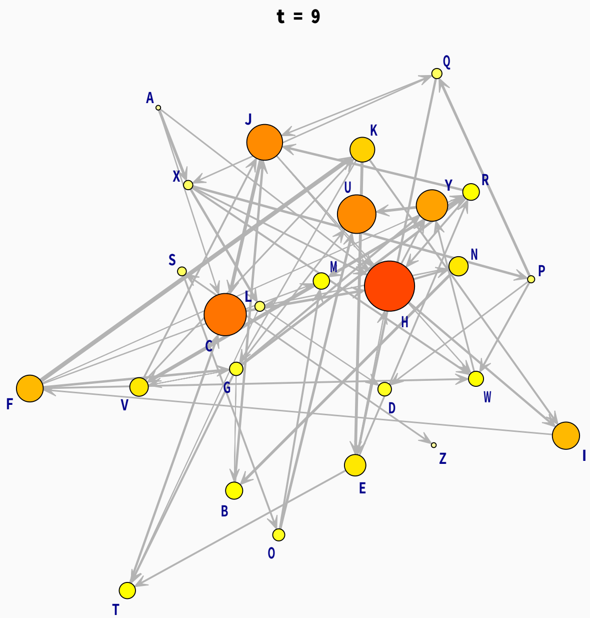
t = 7



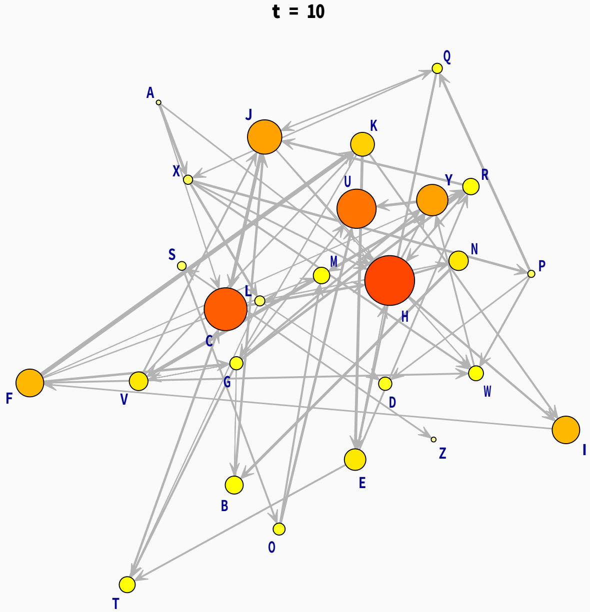
- $\alpha = 0.85$
- initial: 12 nodes



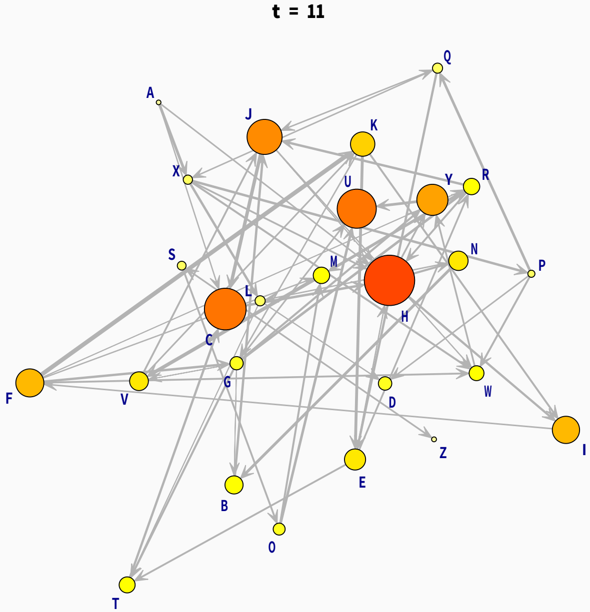
- $\alpha = 0.85$
- initial:
  12 nodes



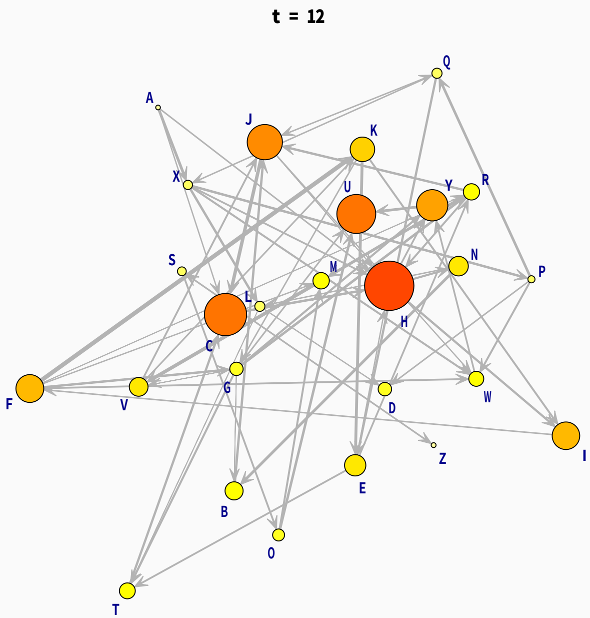
- $\alpha = 0.85$
- initial:
  12 nodes



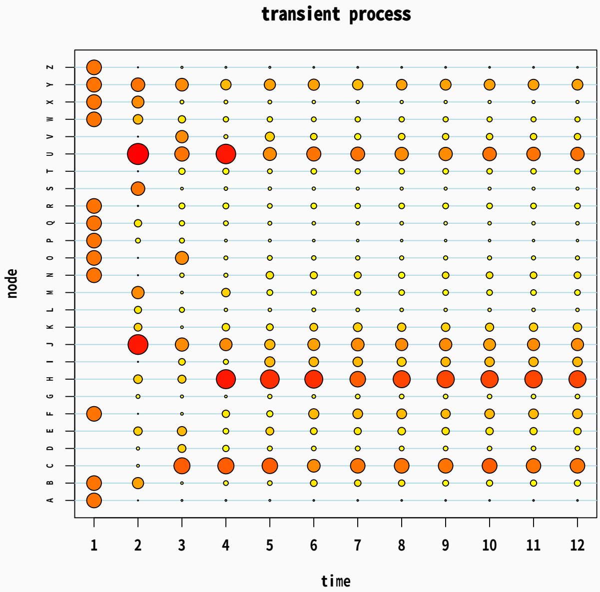
- $\alpha = 0.85$
- initial:
  12 nodes



- $\alpha = 0.85$
- initial:
  12 nodes

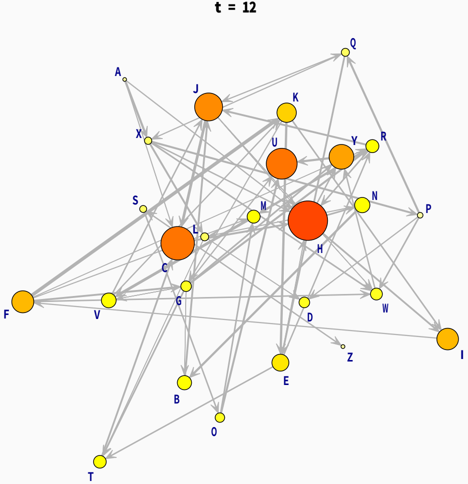


- $\alpha = 0.85$
- initial:
  12 nodes

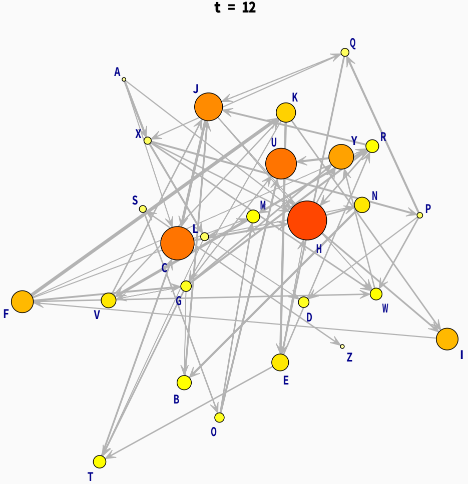


- $\alpha = 0.85$
- initial: 12 nodes

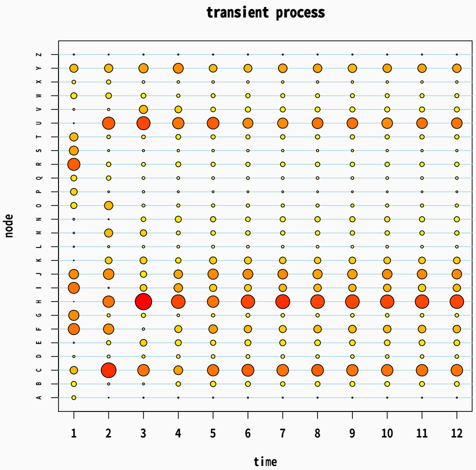
initial condition 1



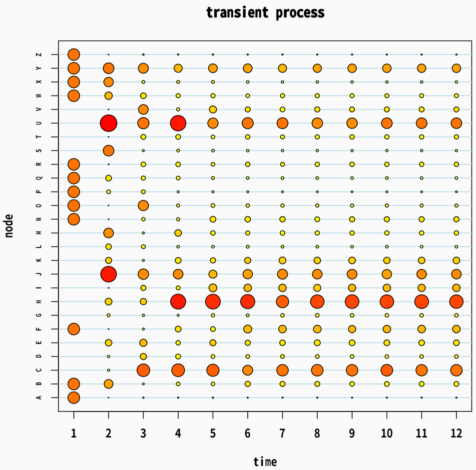
initial condition 2



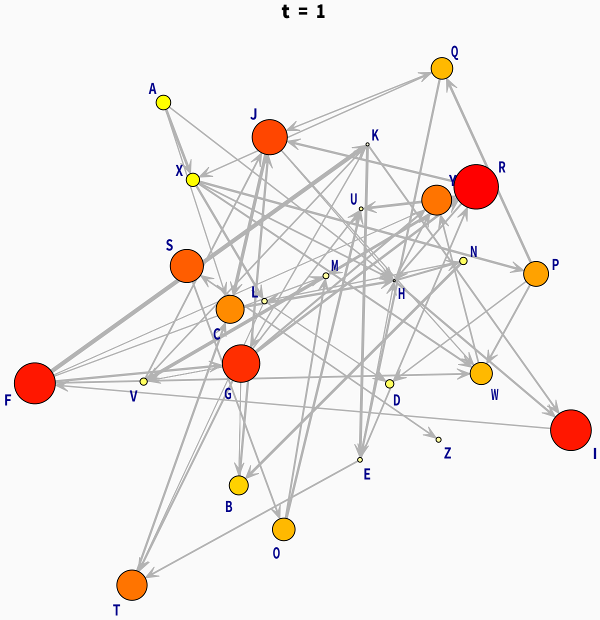
fast convergence to the same stationary distribution initial condition 1



initial condition 2

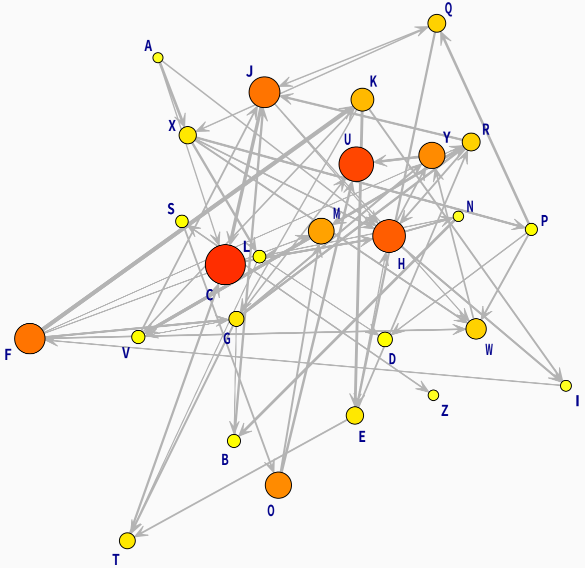


fast convergence to the same stationary distribution

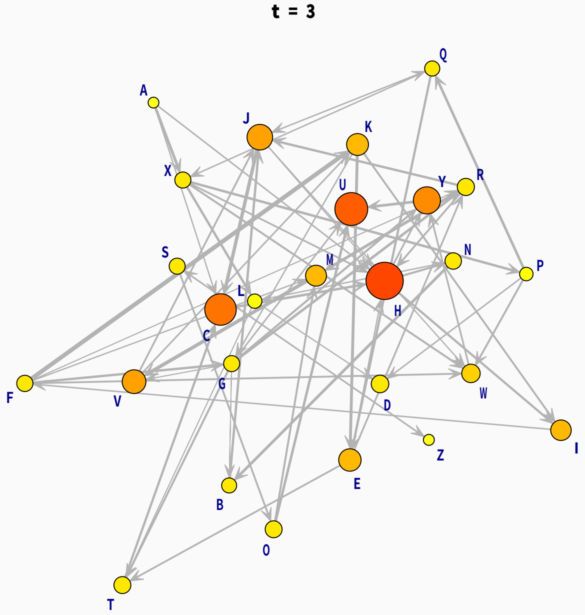


- $\alpha = 0.5$
- initial: uniform dist.

t = 2

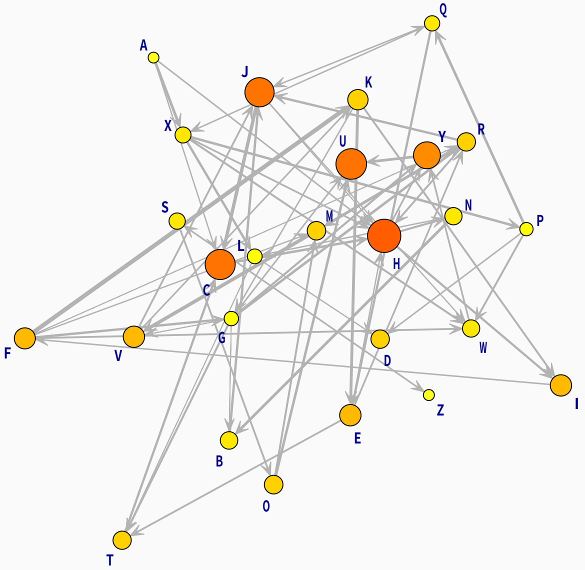


- $\alpha = 0.5$
- initial: uniform dist.

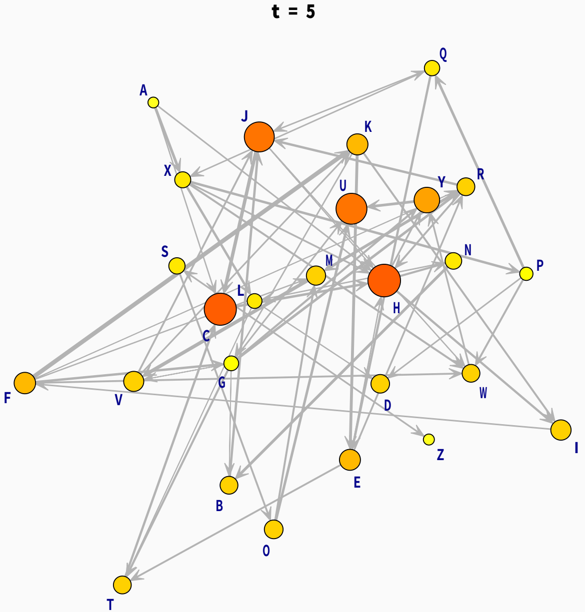


- $\alpha = 0.5$
- initial:
  uniform dist.

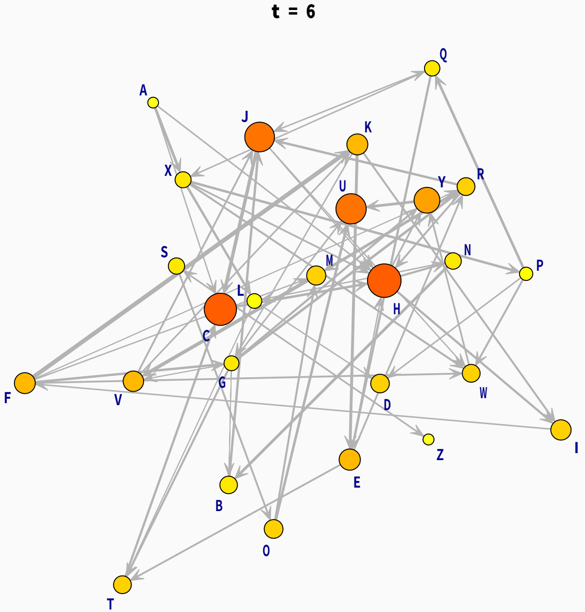
t = 4



- $\alpha = 0.5$
- initial:
  uniform dist.

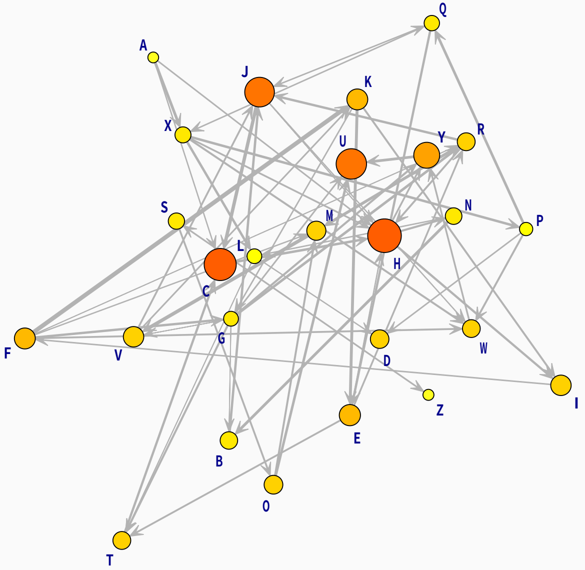


- $\alpha = 0.5$
- initial:
  uniform dist.

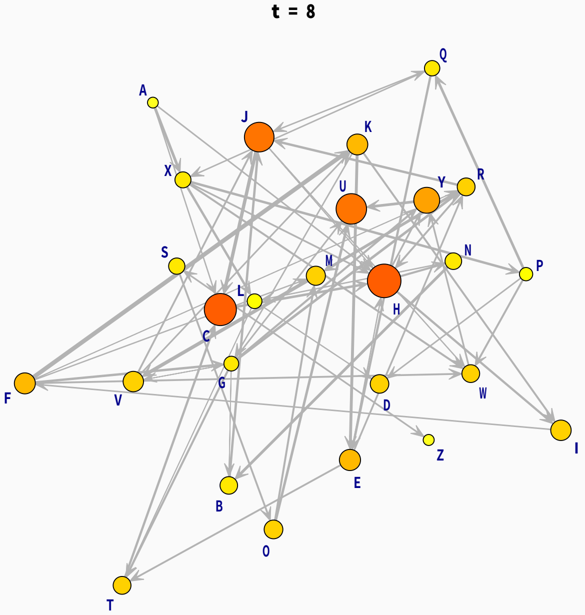


- $\alpha = 0.5$
- initial:
  uniform dist.

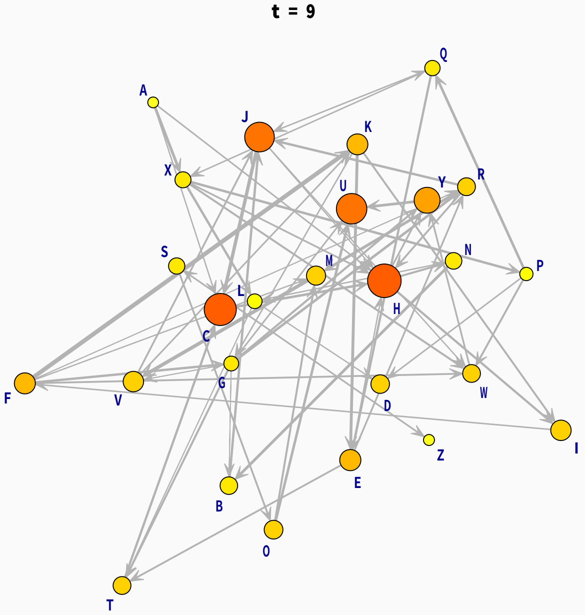
t = 7



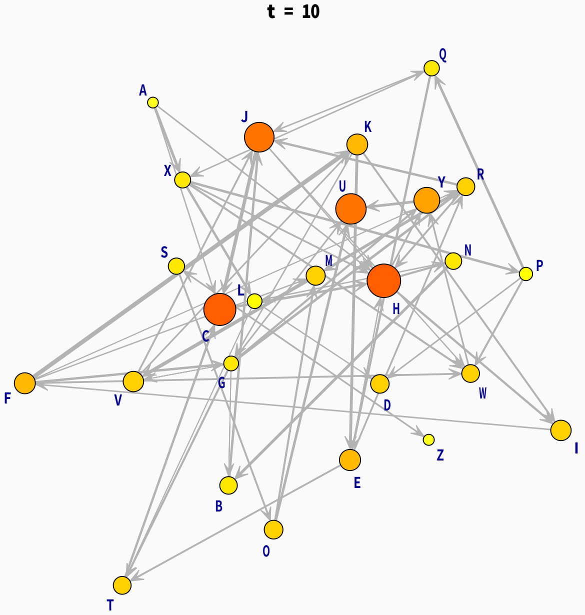
- $\alpha = 0.5$
- initial: uniform dist.



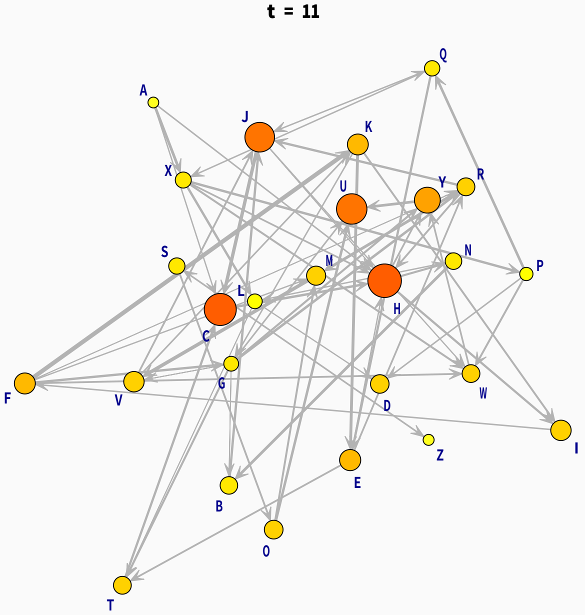
- $\alpha = 0.5$
- initial: uniform dist.



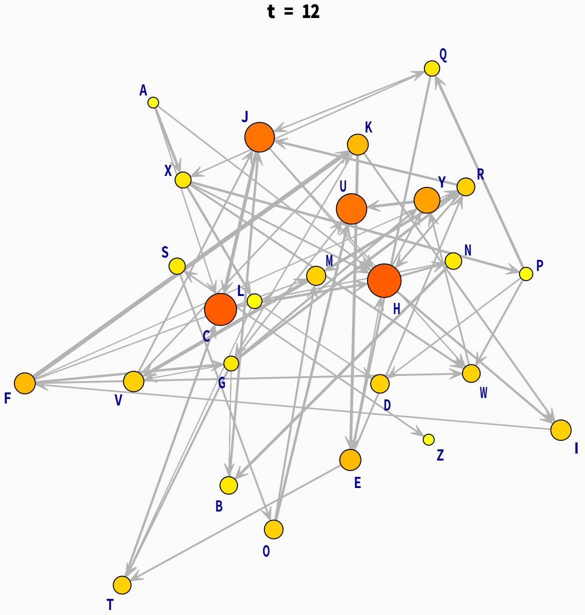
- $\alpha = 0.5$
- initial: uniform dist.



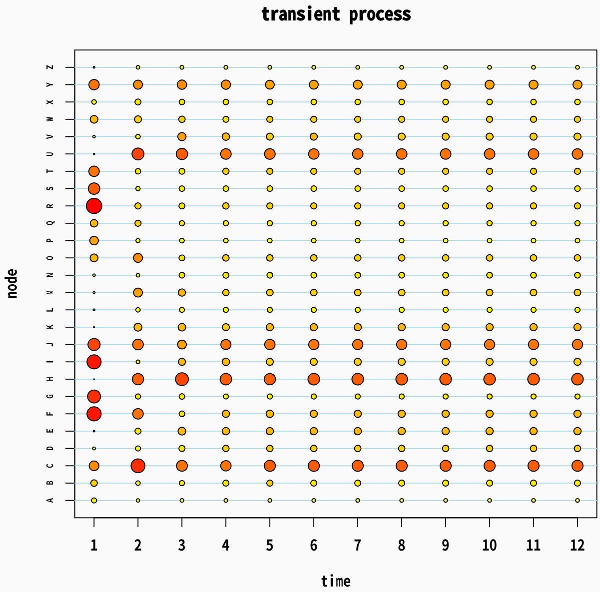
- $\alpha = 0.5$
- initial:
  uniform dist.



- $\alpha = 0.5$
- initial:
  uniform dist.

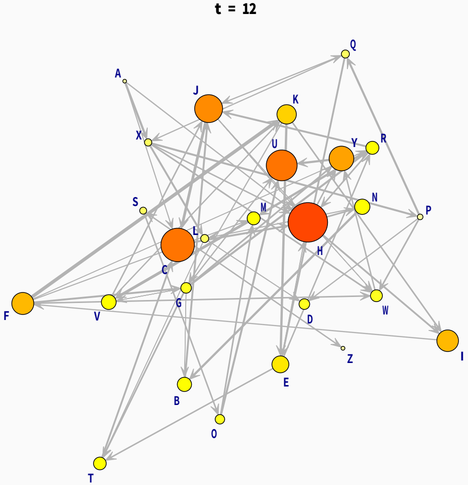


- $\alpha = 0.5$
- initial:
  uniform dist.

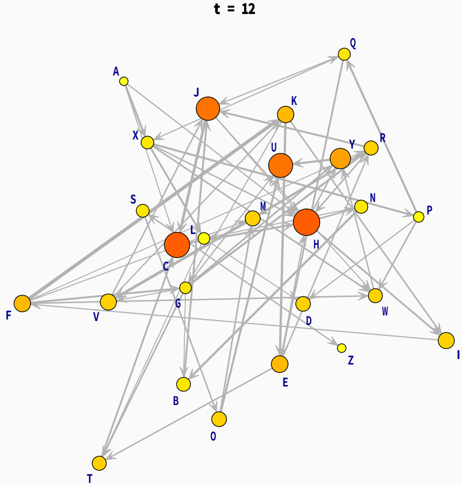


- $\alpha = 0.5$
- initial: uniform dist.

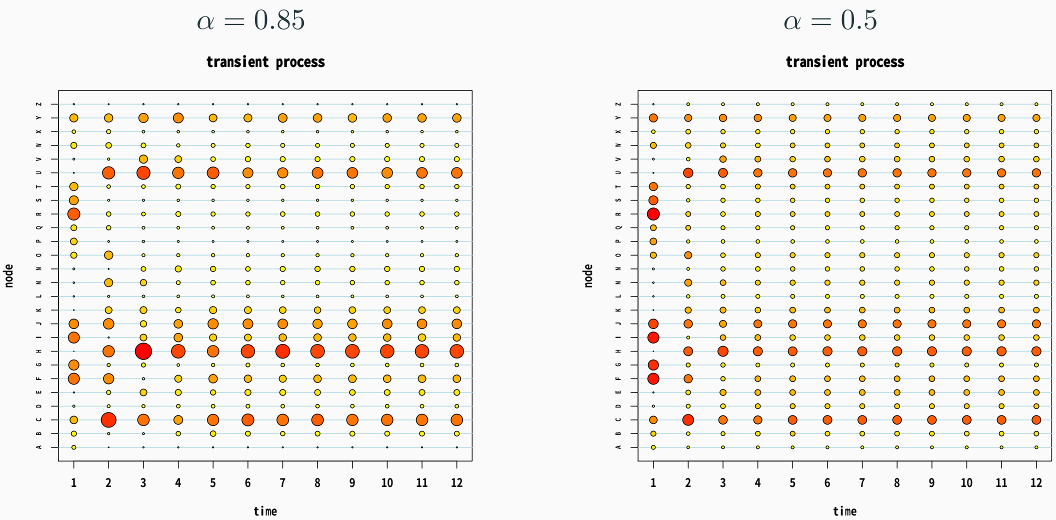
$\alpha = 0.85$



$\alpha = 0.5$

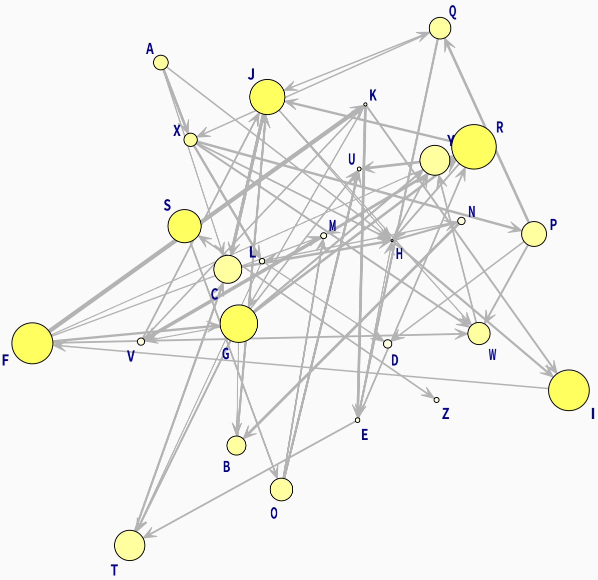


random diffusion with smaller scaling

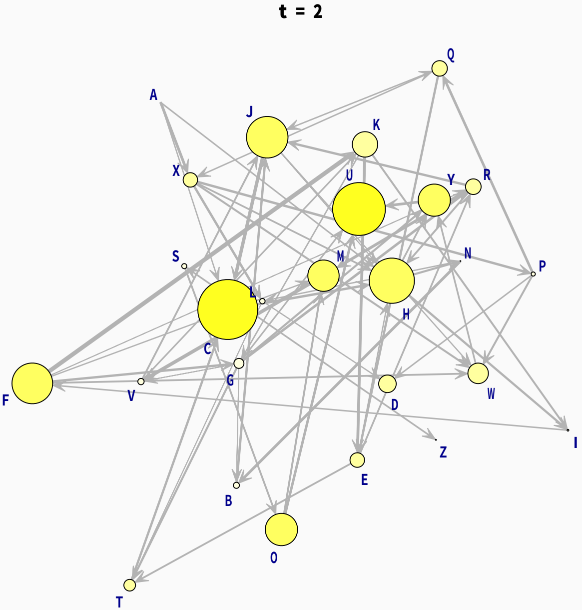


random diffusion with smaller scaling

t = 1

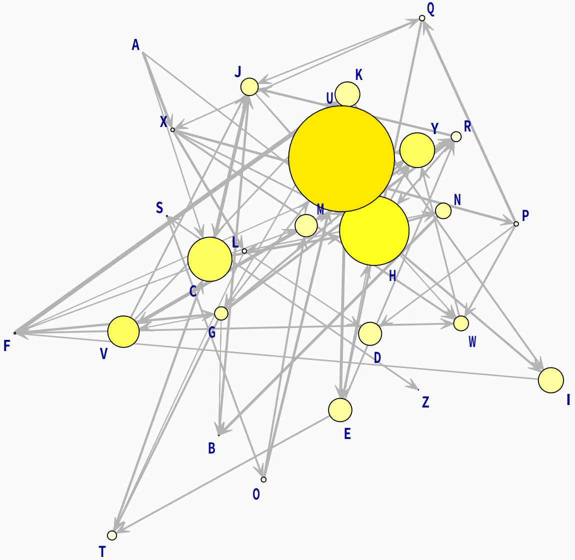


- without sink node escape
- initial:
  uniform dist.

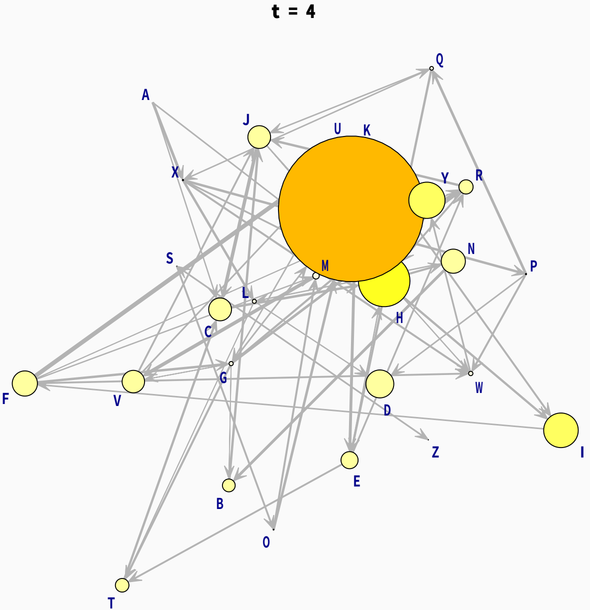


- without sink node escape
- initial:
  uniform dist.

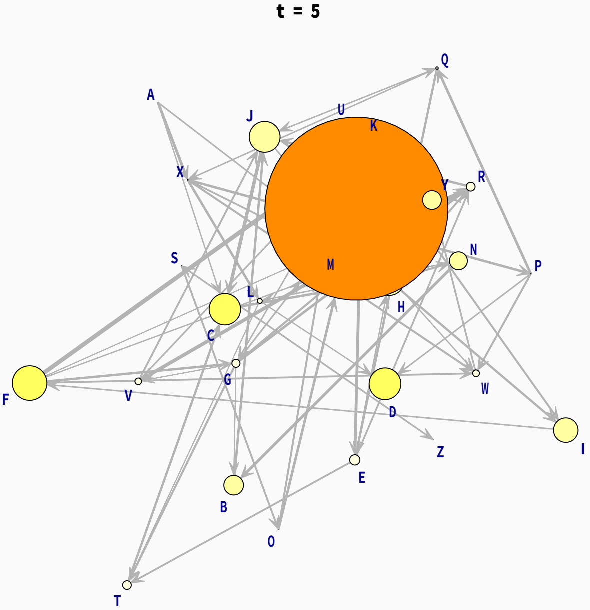
t = 3



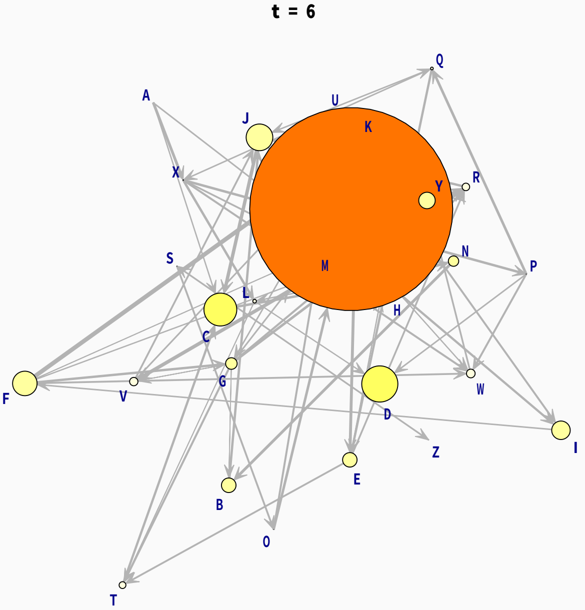
- without sink node escape
- initial:
  uniform dist.



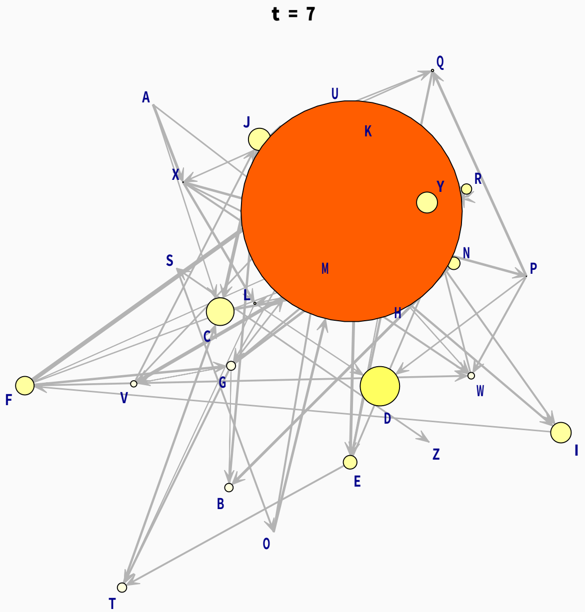
- without sink node escape
- initial:
  uniform dist.



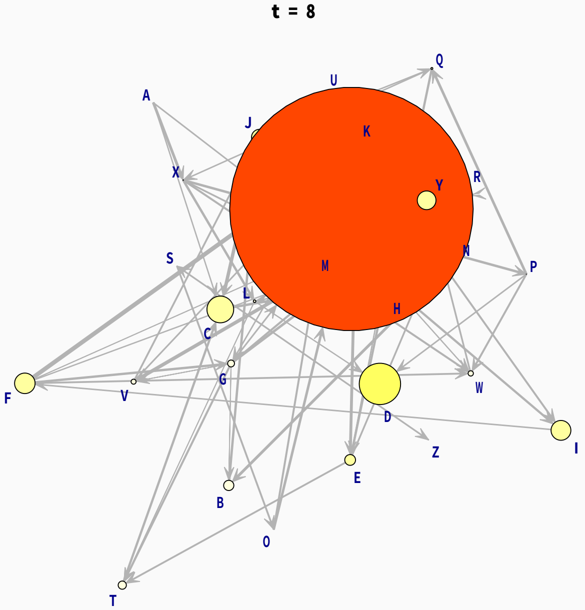
- without sink node escape
- initial:
  uniform dist.



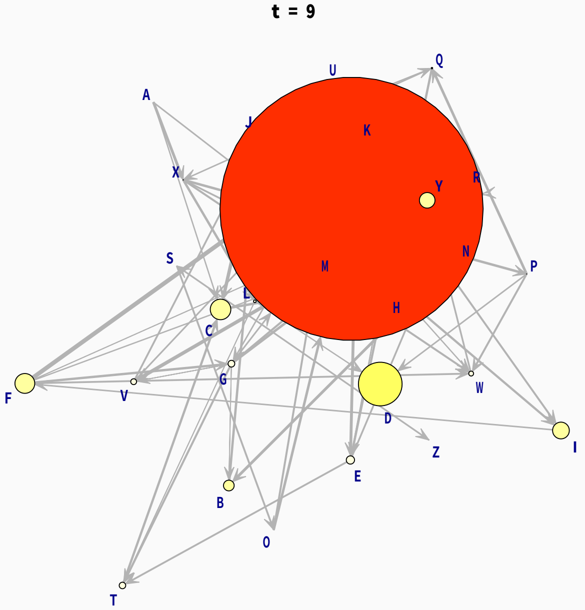
- without sink node escape
- initial:
  uniform dist.



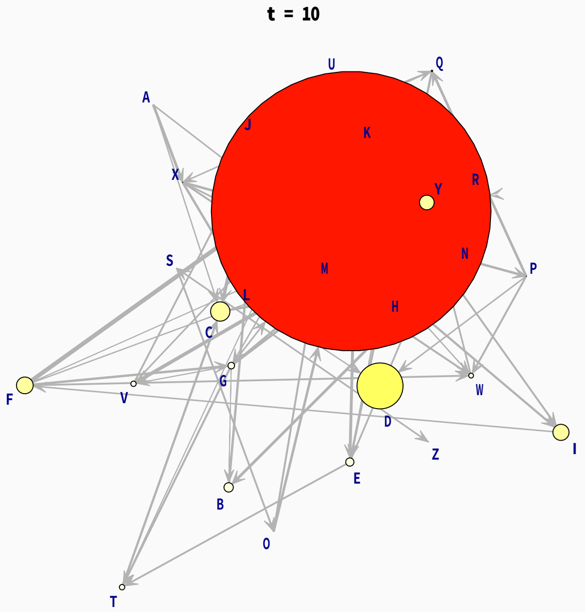
- without sink node escape
- initial:
  uniform dist.



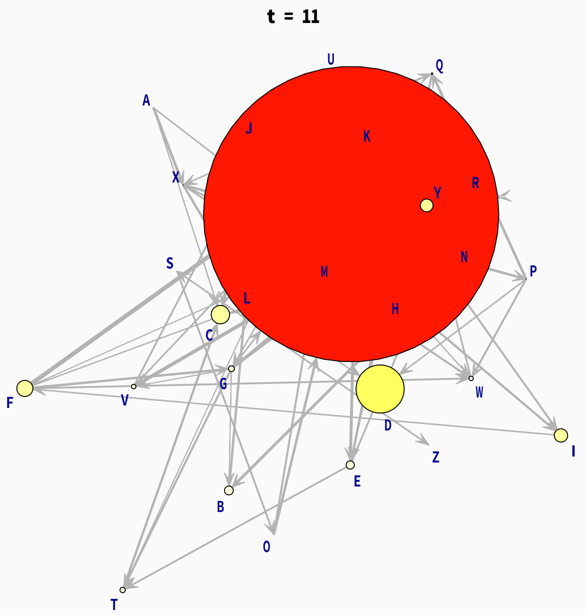
- without sink node escape
- initial:
  uniform dist.



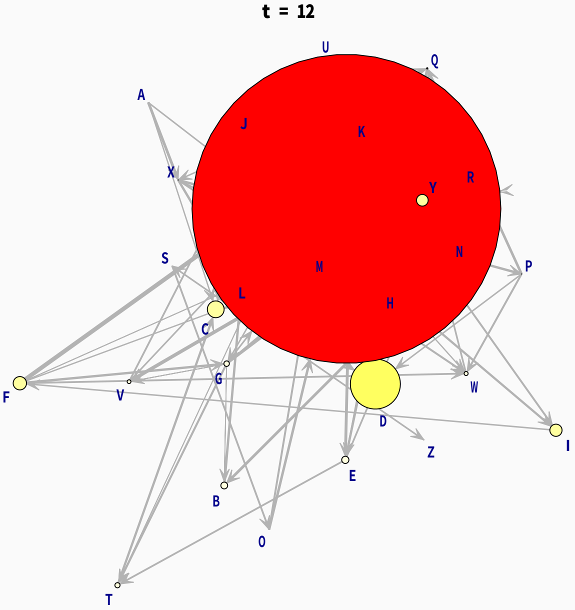
- without sink node escape
- initial:
  uniform dist.



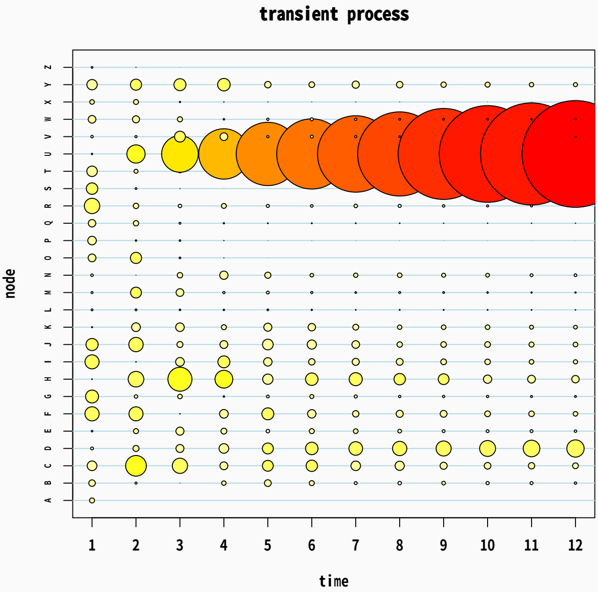
- without sink node escape
- initial:
  uniform dist.



- without sink node escape
- initial:
  uniform dist.

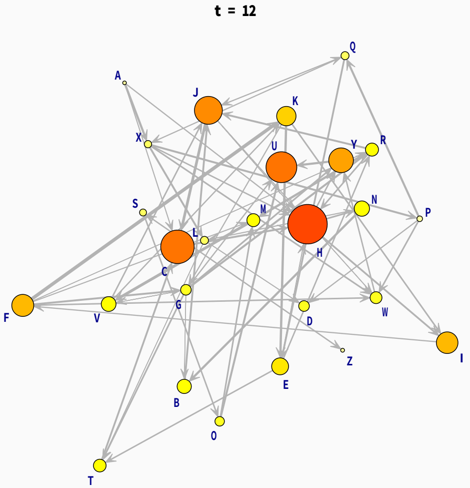


- without sink node escape
- initial:
  uniform dist.

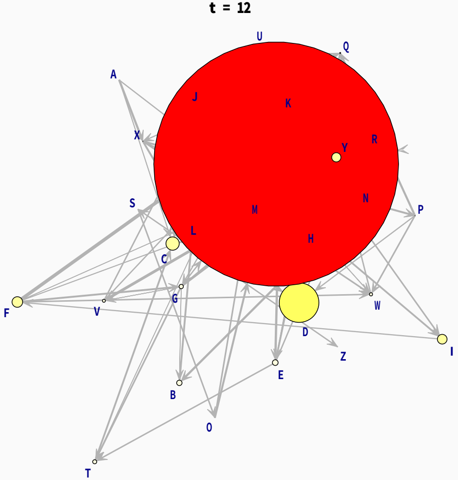


- without sink node escape
- initial:
  uniform dist.

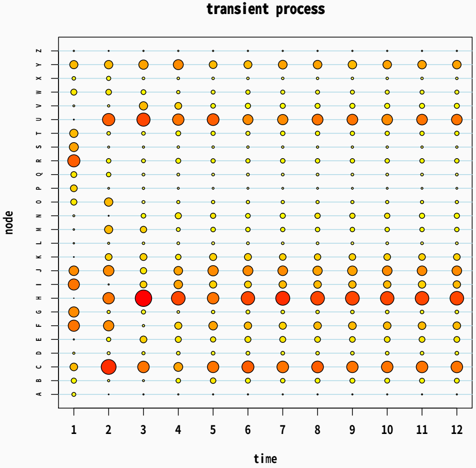
with sink node adjustment



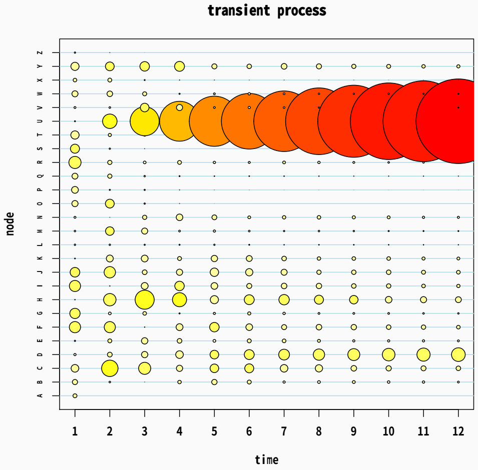
without sink node adjustment



concentration on a specific node ("sink") with sink node adjustment



without sink node adjustment



concentration on a specific node (“sink”) a simple and strong model of movements on directed graph:

- behavior model of selection from finite options
  - web surf model
  - purchase model of certain genres
  - transition model of audience ratings
  - customer share model of restaurants/coffee shops
- adjustment of transition matrix (sink node/alpha)
  - out of stock or service
  - capricious or adventurous attempts
  - introduction from others

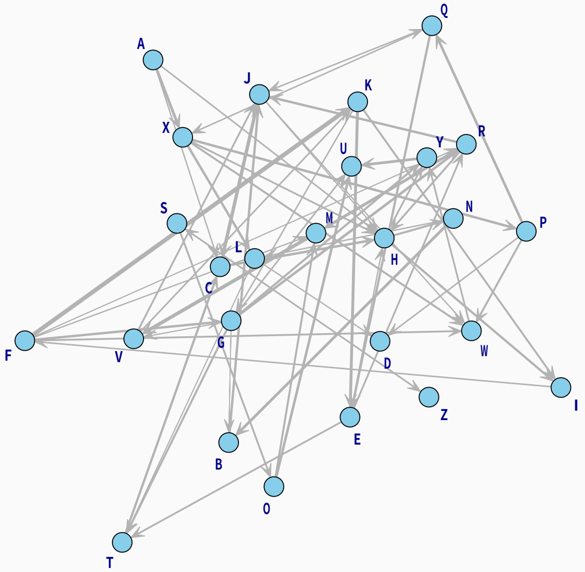
Problem Formulation

non-stationary data on directed graphs:

- strength of edges slowly change in time
  - change of structure
  - change of stationary distribution
- model assumption:
  - frequent update (fast time scale; t)
    e.g.: purchase every day
  - sparse observation (slow time scale; T)
    e.g.: aggregate every week

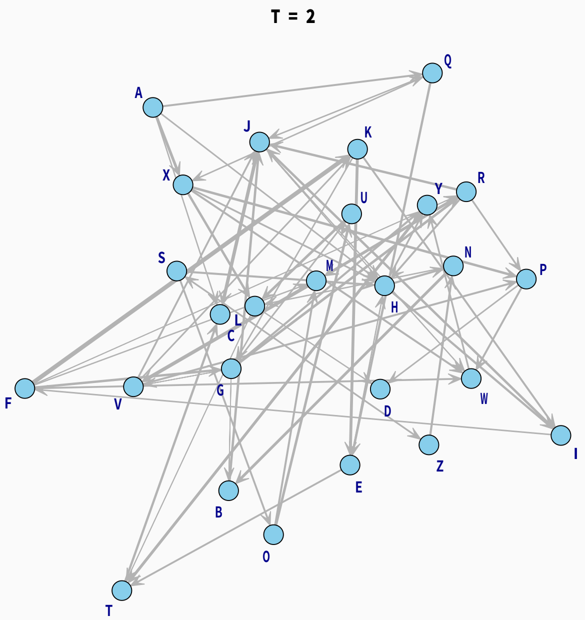
  observations are supposed to be on stationary distribution at current point

T = 1



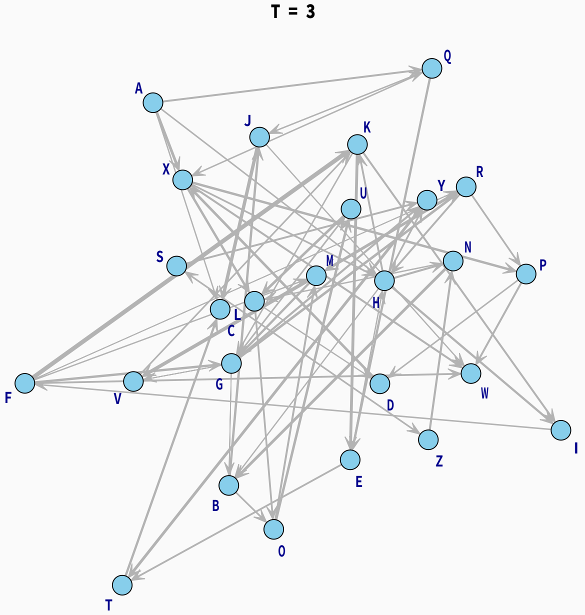
change of structure

- 10% edges at random
- relatively small



change of structure

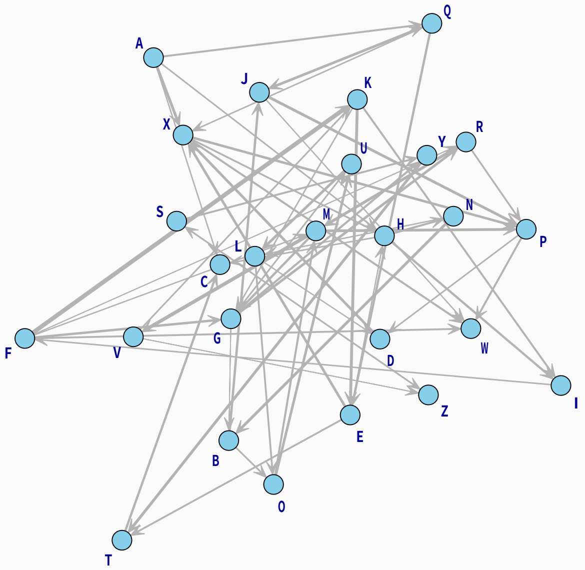
- 10% edges at random
- relatively small



change of structure

- 10% edges at random
- relatively small

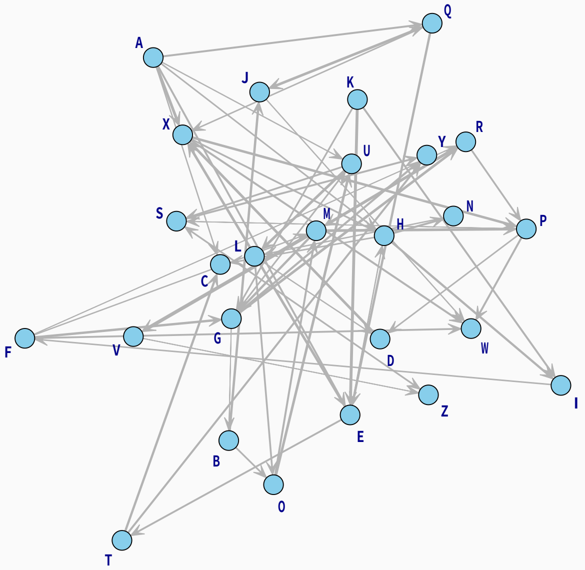
T = 4



change of structure

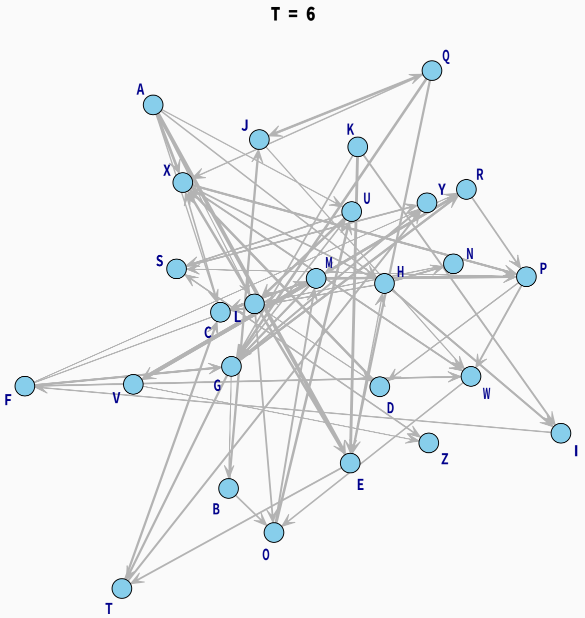
- 10% edges at random
- relatively small

T = 5



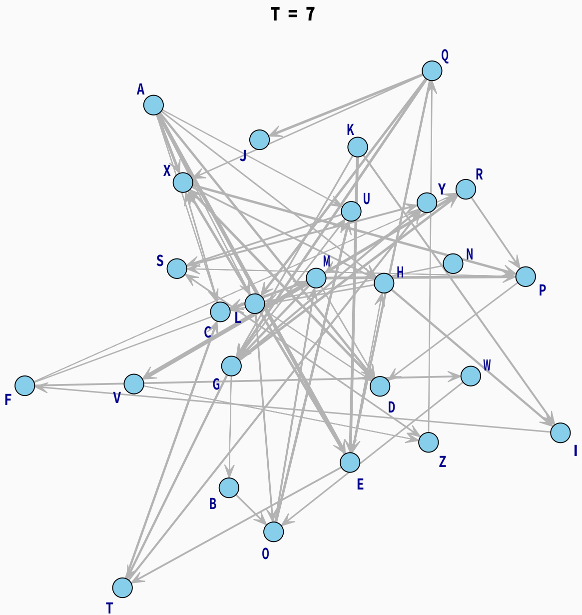
change of structure

- 10% edges at random
- relatively small



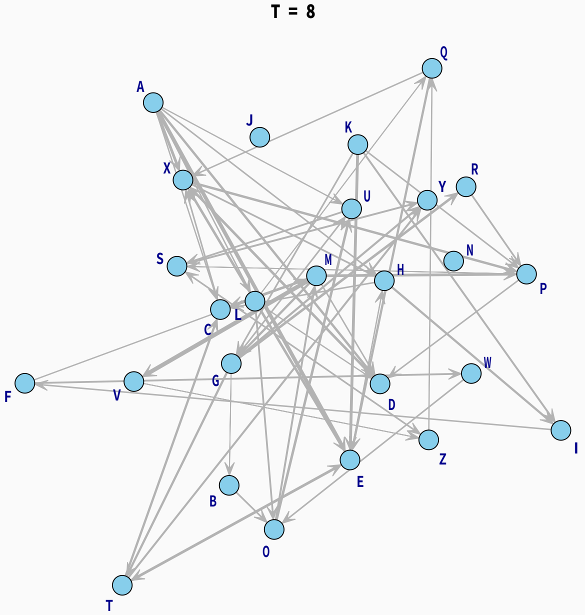
change of structure

- 10% edges at random
- relatively small



change of structure

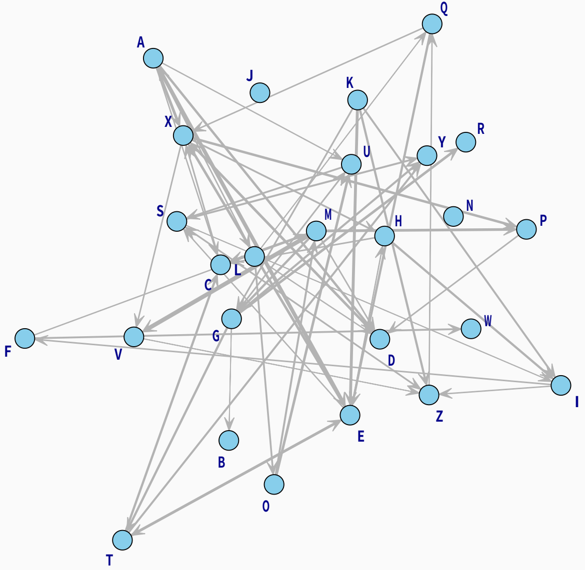
- 10% edges at random
- relatively small



change of structure

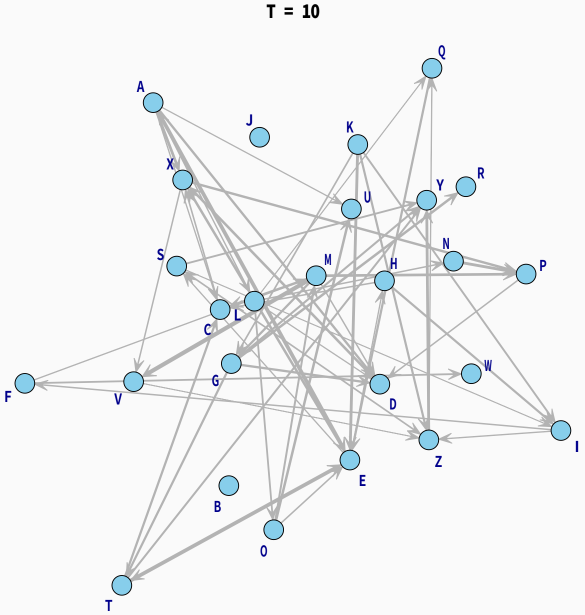
- 10% edges at random
- relatively small

T = 9



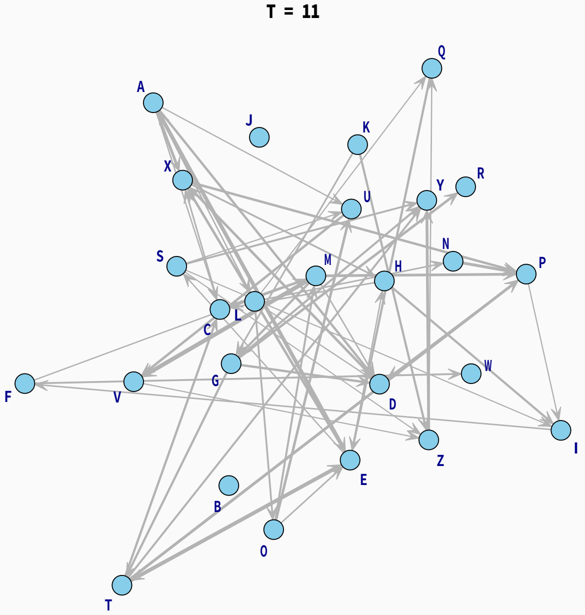
change of structure

- 10% edges at random
- relatively small



change of structure

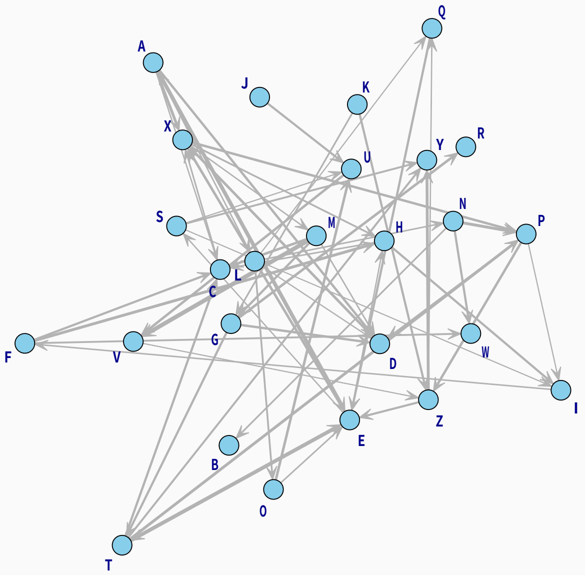
- 10% edges at random
- relatively small



change of structure

- 10% edges at random
- relatively small

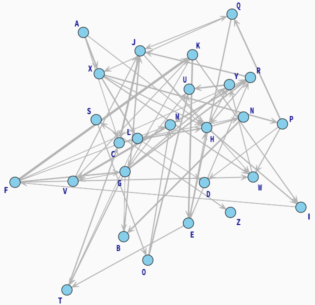
T = 12



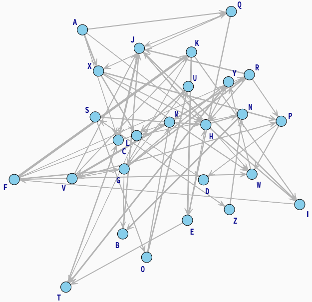
change of structure

- 10% edges at random
- relatively small

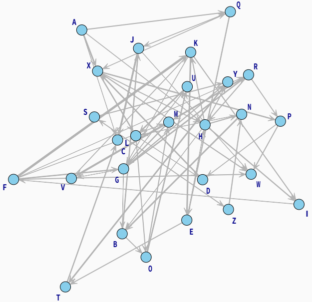
T = 1



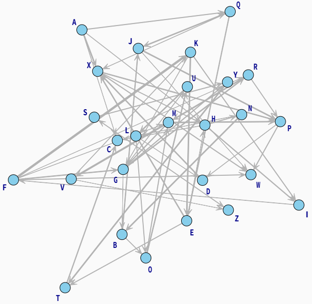
T = 2



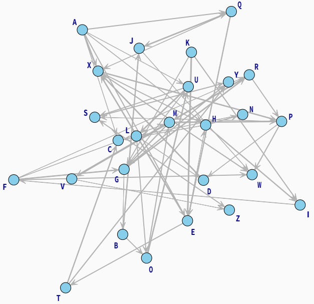
T = 3



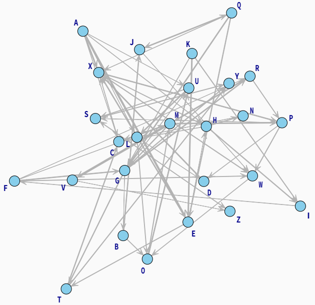
T = 4

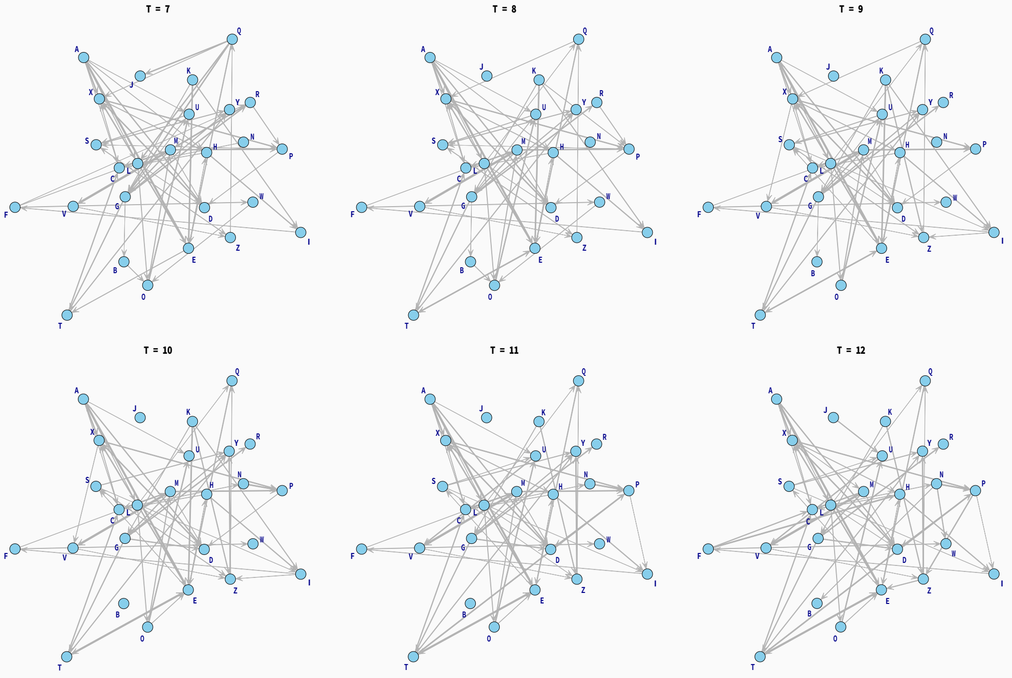


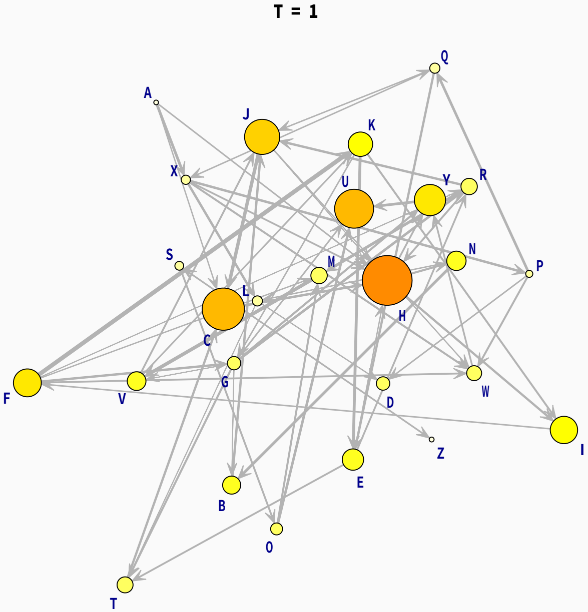
T = 5



T = 6

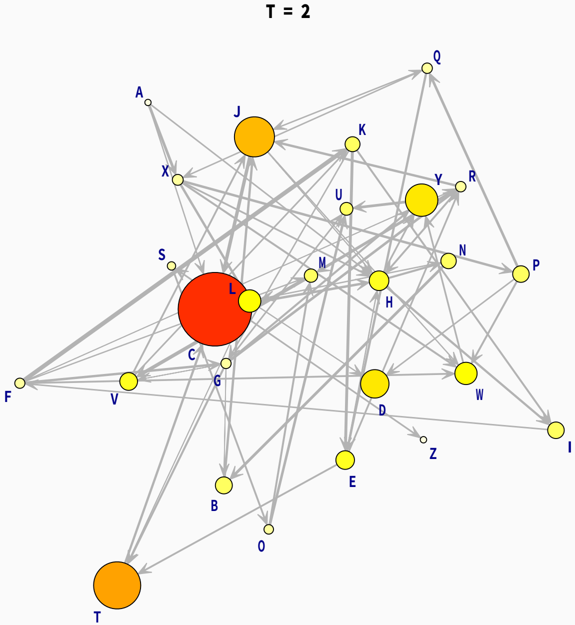



T = 7
T = 8
T = 9
T = 10
T = 11
T = 12



change of stationary dist.%ribution

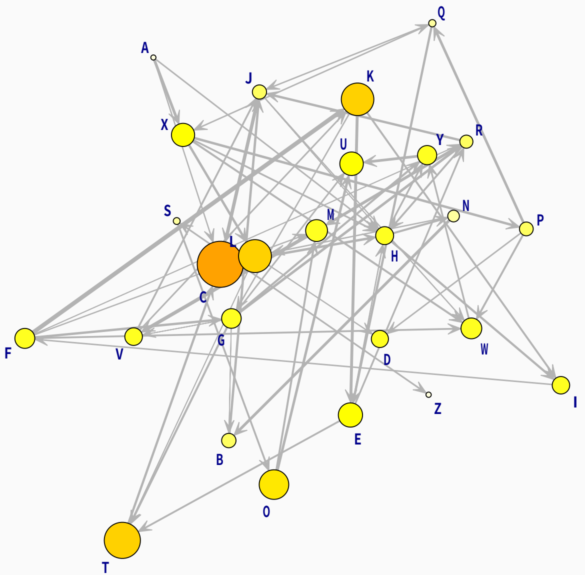
- 10% edges at random
- large effect



change of stationary dist.%ribution

- 10% edges at random
- large effect

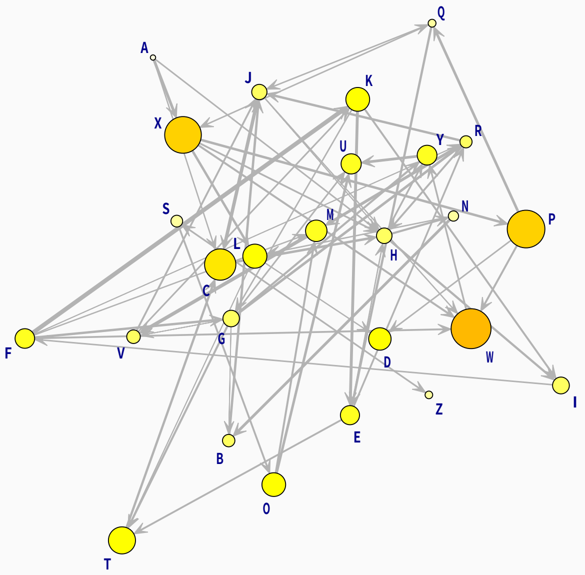
T = 3



change of stationary dist.%ribution

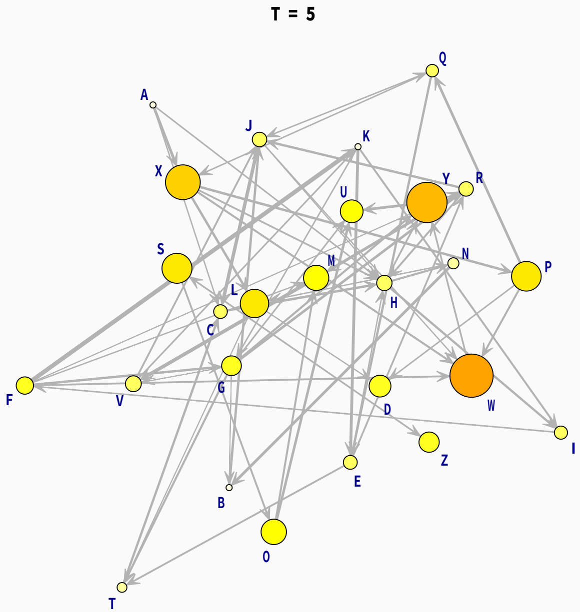
- 10% edges at random
- large effect

T = 4



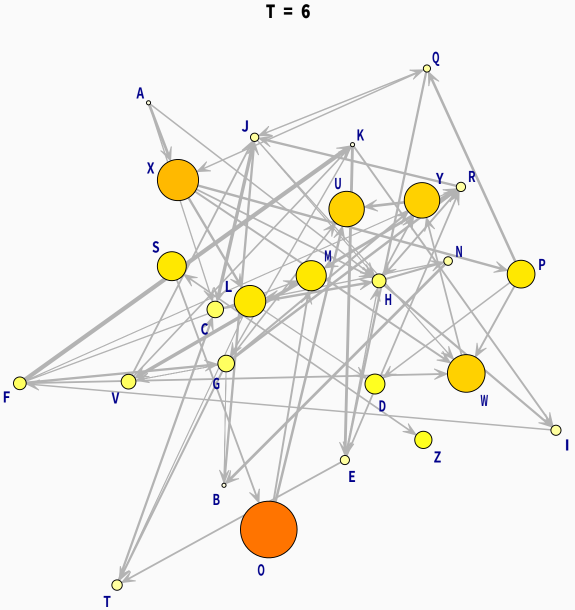
change of stationary dist.%ribution

- 10% edges at random
- large effect



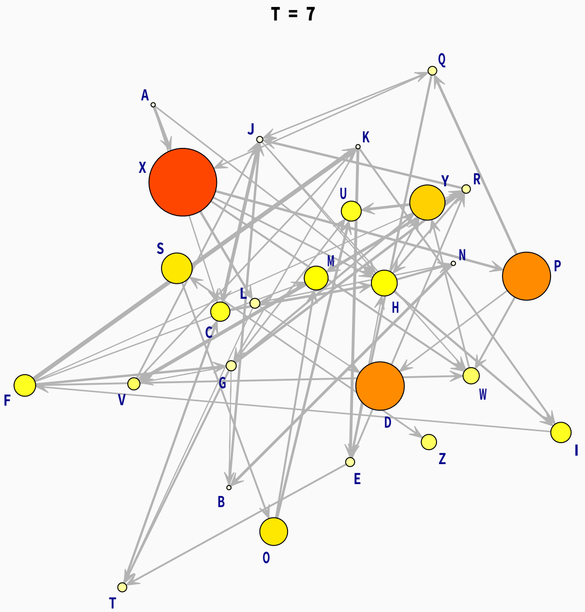
change of stationary dist.%ribution

- 10% edges at random
- large effect



change of stationary dist.%ribution

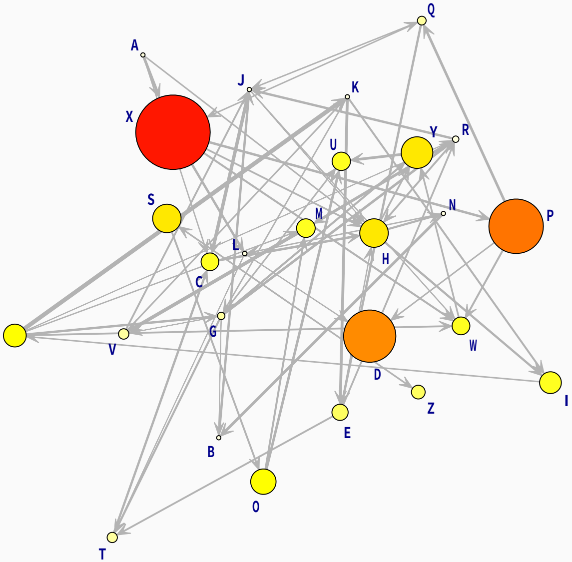
- 10% edges at random
- large effect



change of stationary dist.%ribution

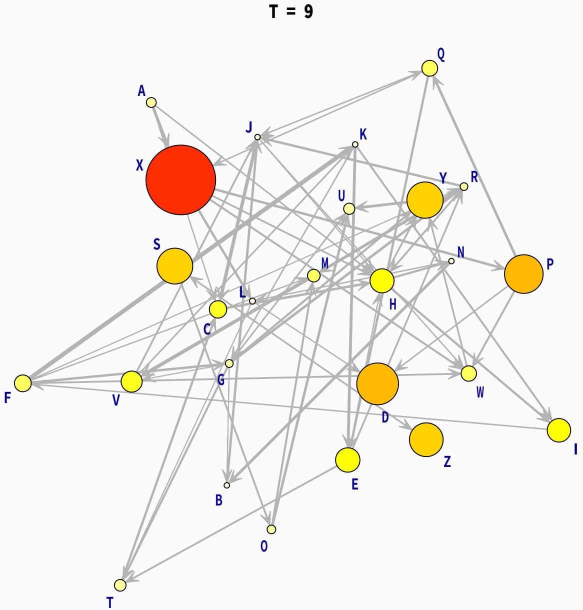
- 10% edges at random
- large effect

T = 8



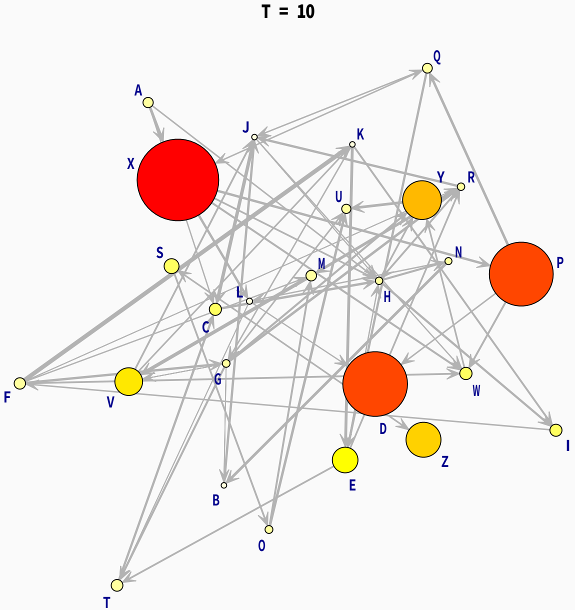
change of stationary dist.%ribution

- 10% edges at random
- large effect



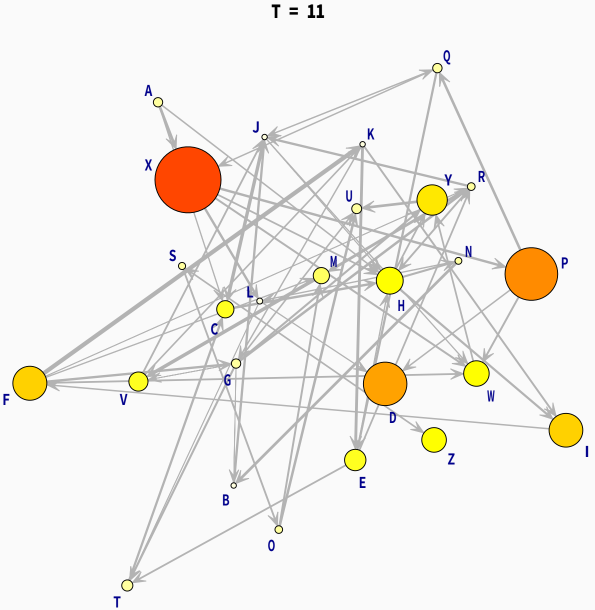
change of stationary dist.%ribution

- 10% edges at random
- large effect



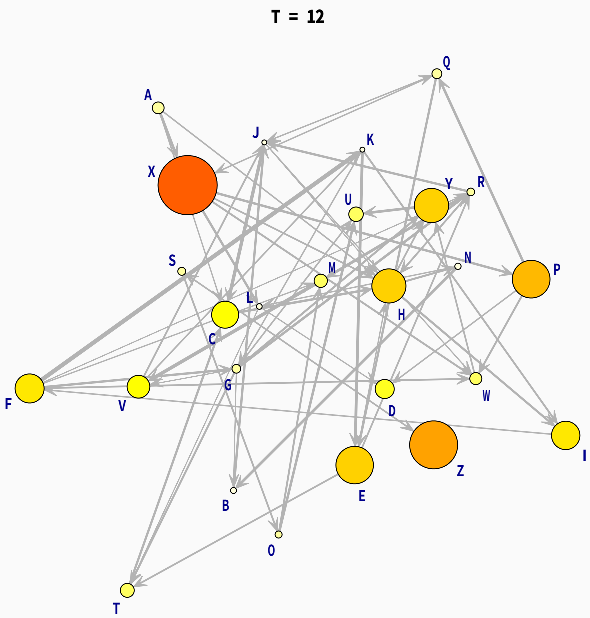
change of stationary dist.%ribution

- 10% edges at random
- large effect



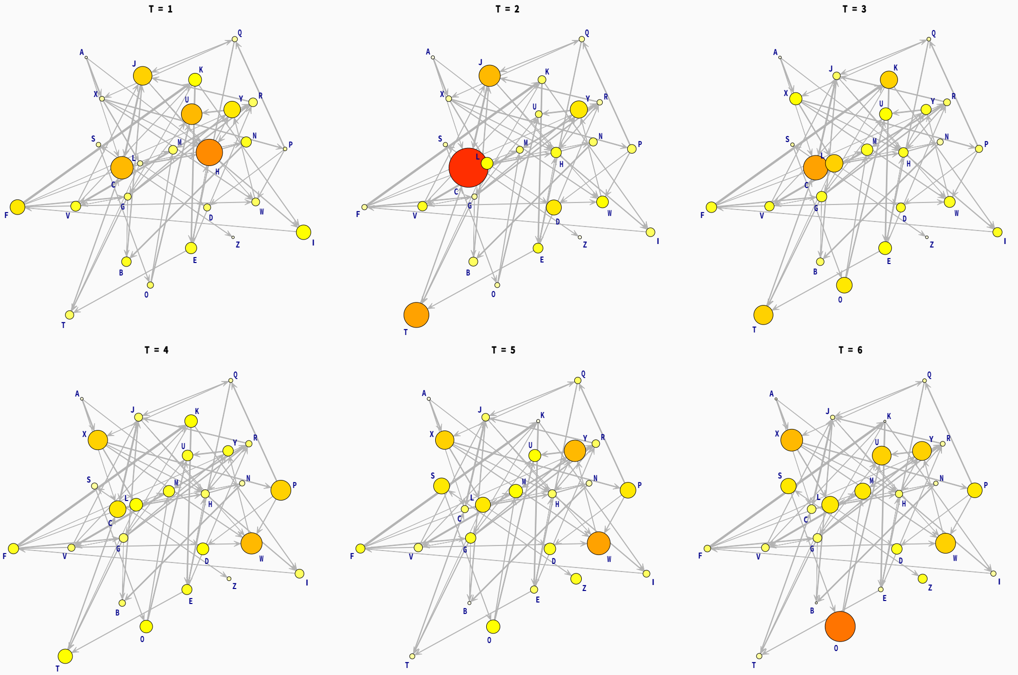
change of stationary dist.%ribution

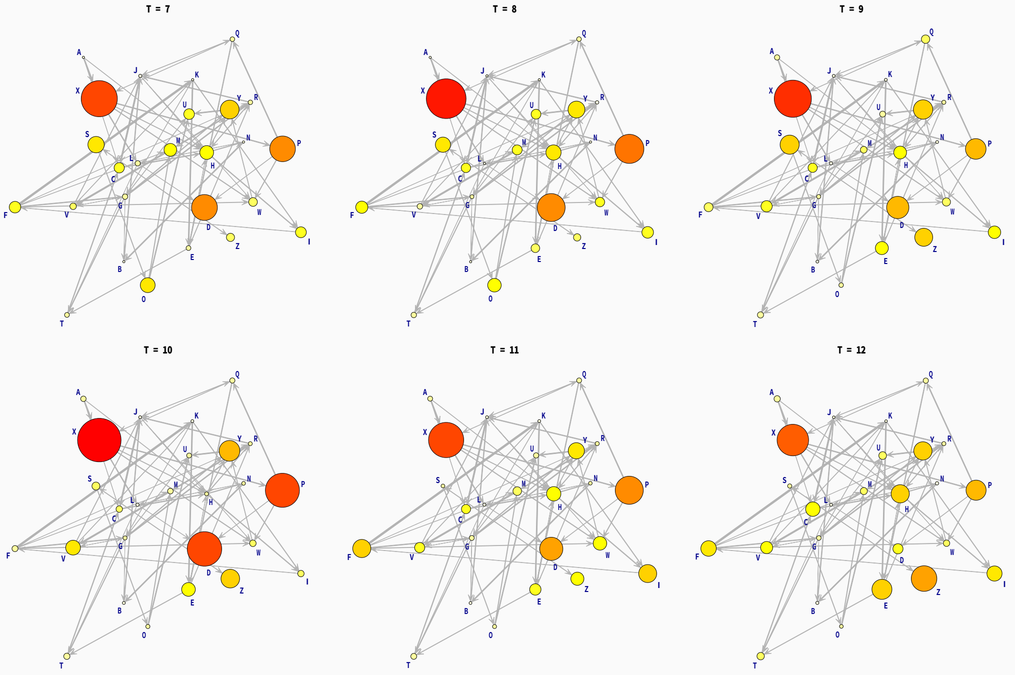
- 10% edges at random
- large effect

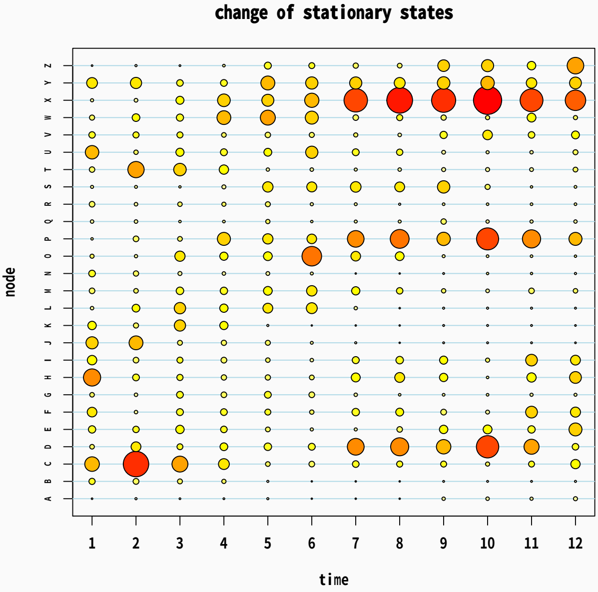


change of stationary dist.%ribution

- 10% edges at random
- large effect


T = 1
T = 2
T = 3
T = 4
T = 5
T = 6


T = 7
T = 8
T = 9
T = 10
T = 11
T = 12



change of statinary dist.%ribution

- 10% edges at random
- large effect

Problem

for given series of stationary distributions $\{\pi_t\}$, estimate series of graph structures $\{G_t\}$.

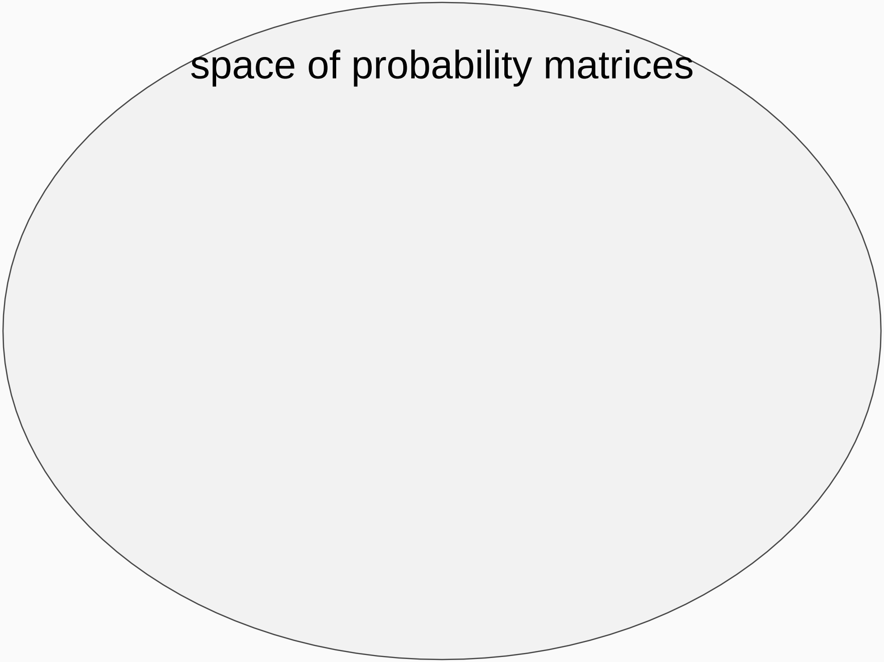
$$\text{minimize } L(\{G_t\}) \quad \text{subject to } \forall t,\ \pi_t^T G_t = \pi_t^T$$

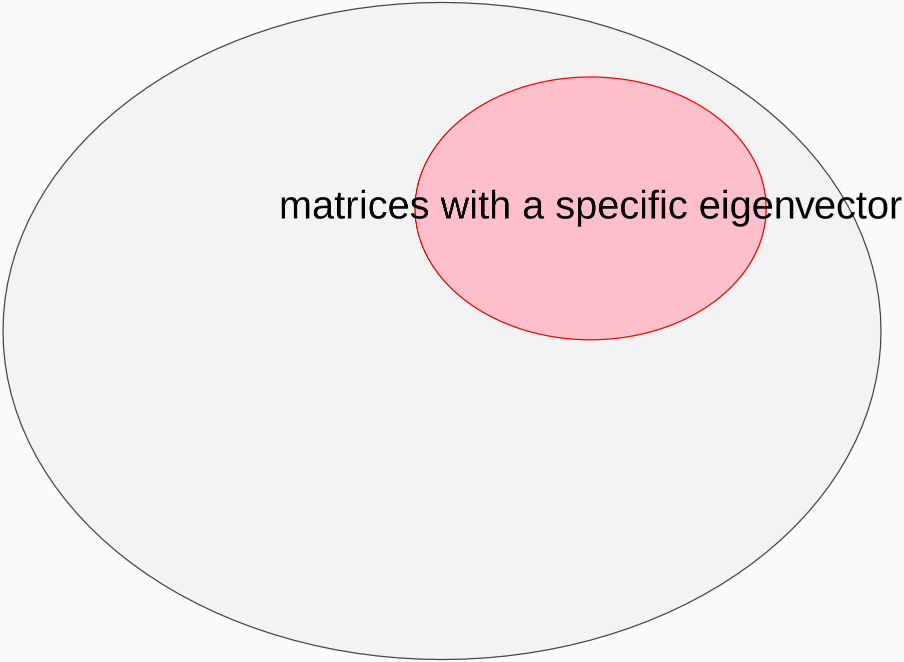
difficulties:

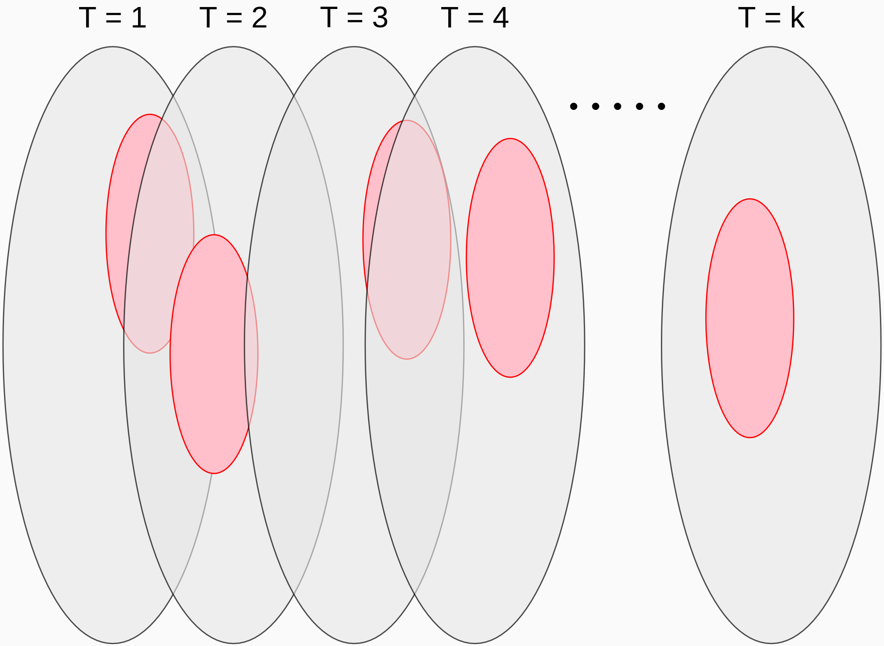
- infinitely many matrices have the same eigenvector
- the followings are needed:
  - assumptions on graph structures
  - assumptions of graph changes

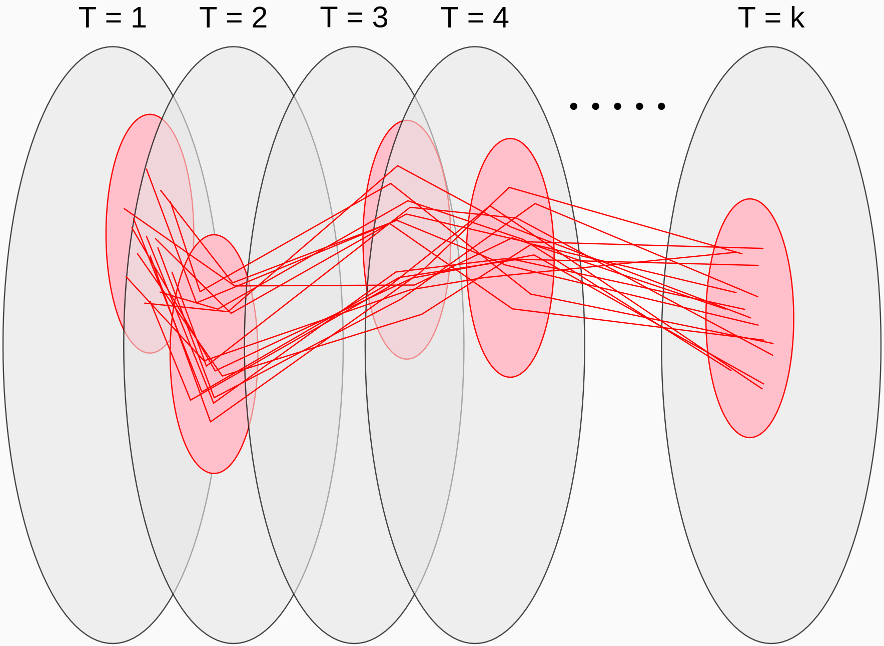
e.g. (fused lasso): for a certain sparse norm of matrix, $\|\cdot\|_s$,

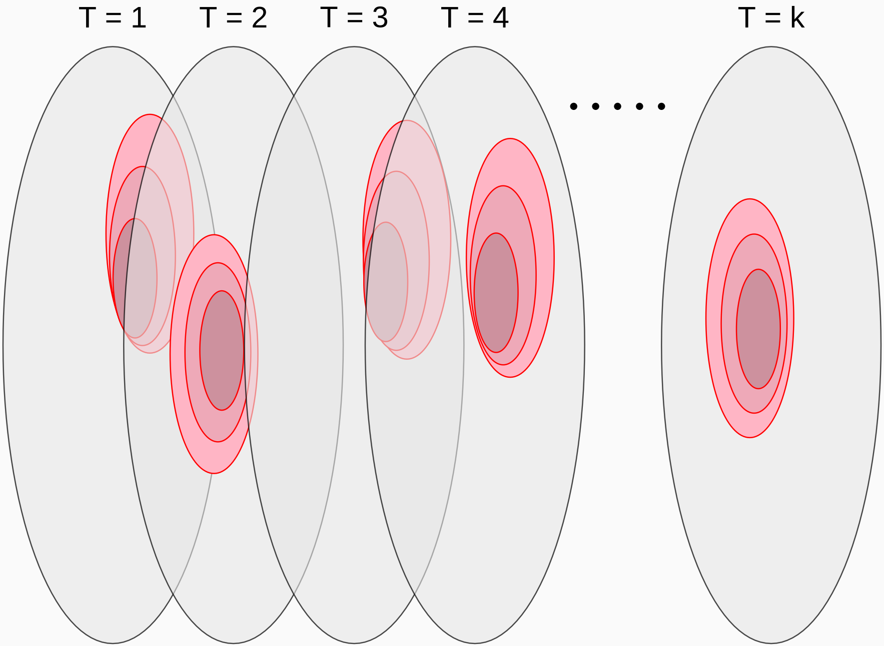
$$L(\{G_t\}) = \sum_t \|G_t\|_s + \sum_t \|G_{t+1} - G_t\|_s$$

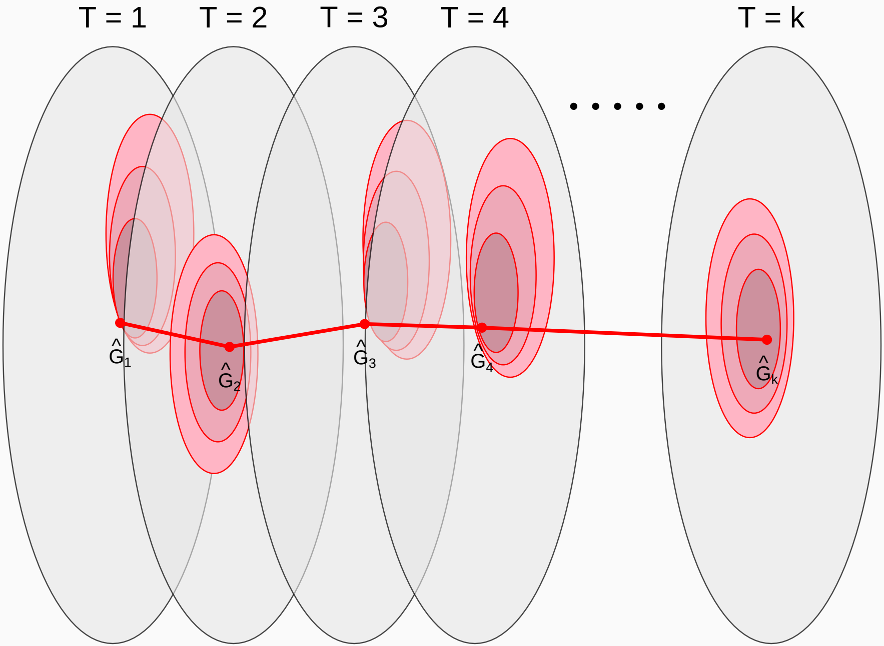

space of probability matrices


matrices with a specific eigenvector


T = 1
T = 2
T = 3
T = 4
T = k


T = 1
T = 2
T = 3
T = 4
T = k


T = 1
T = 2
T = 3
T = 4
T = k


T = 1
T = 2
T = 3
T = 4
T = k
. . . . .
$\hat{G}_1$
$\hat{G}_2$
$\hat{G}_3$
$\hat{G}_4$
$\hat{G}_k$

make matrices keep in a restricted subspace:

- eigenvector of matrix A to be $\boldsymbol{a}$ (unit vector):

$$A\boldsymbol{a} = \lambda\boldsymbol{a} + \boldsymbol{b} \Rightarrow A - \boldsymbol{b}\boldsymbol{a}^T \to A$$

- eigenvalue of eigenvector $\boldsymbol{a}$ to be μ:

$$A\boldsymbol{a} = \lambda\boldsymbol{a} \Rightarrow A + (\mu - \lambda)\boldsymbol{a}\boldsymbol{a}^T \to A$$

- A to be a probability matrix:

$$A\boldsymbol{e} = \boldsymbol{d} \Rightarrow \mathrm{diag}(\boldsymbol{d})^{-1}A \to A$$

Numerical Examples

Reference

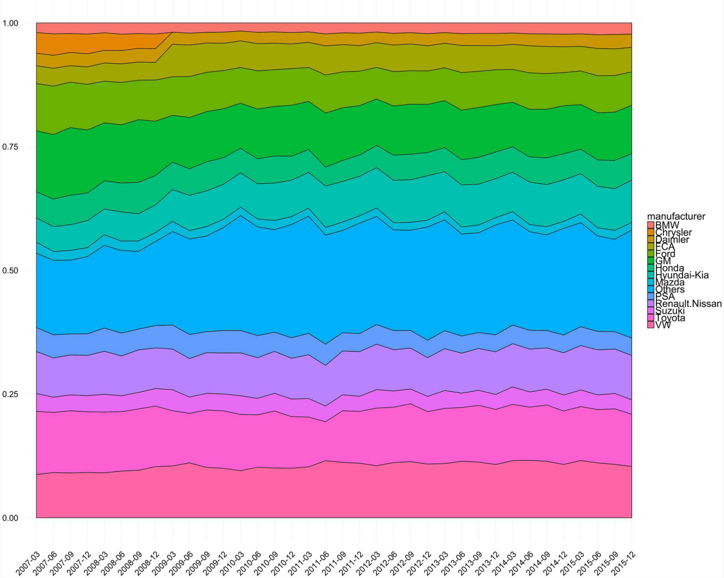
Chiba et al. “Time-Varying Transition Probability Matrix Estimation and Its Application to Brand Share Analysis”

- quarterly unit automobile sales data of manufacturers from 2007-1Q to 2015-4Q
- estimate transition paths and discuss the relation between social events and estimated results
- objective:

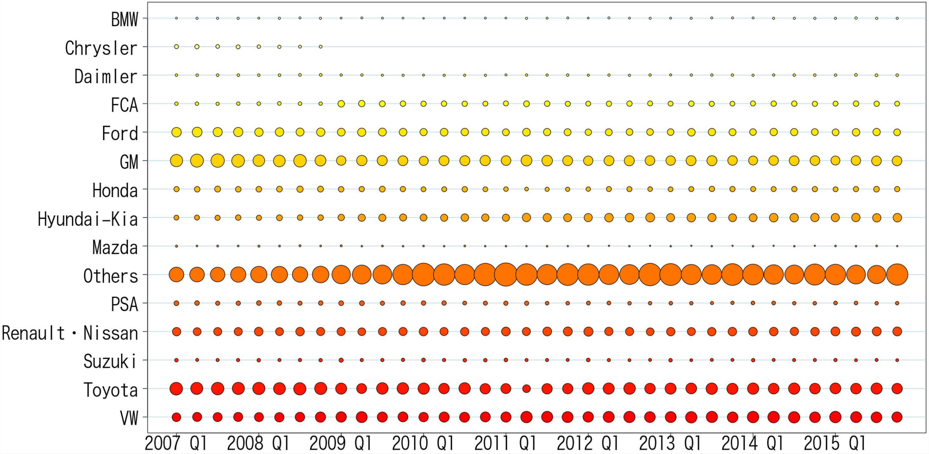
$$L(\{G_t\}) = \sum_t \|G_{t+1} - G_t\|_1$$

- optimization: simplex method with slack variables

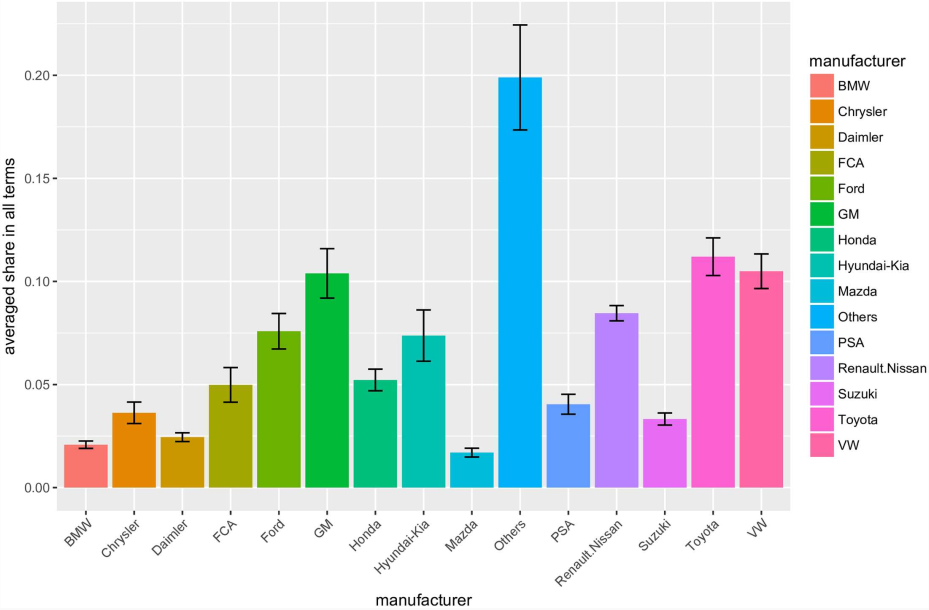
automobile sales for different manufactures



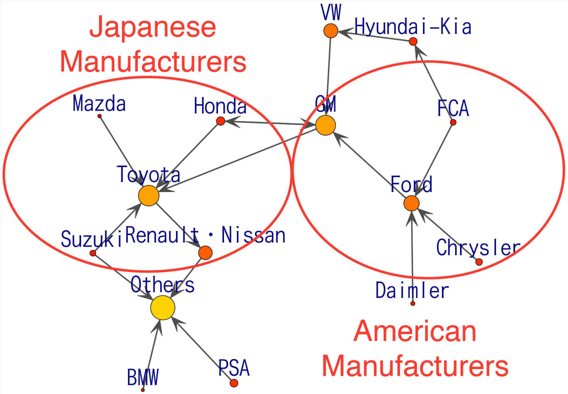
market share transition



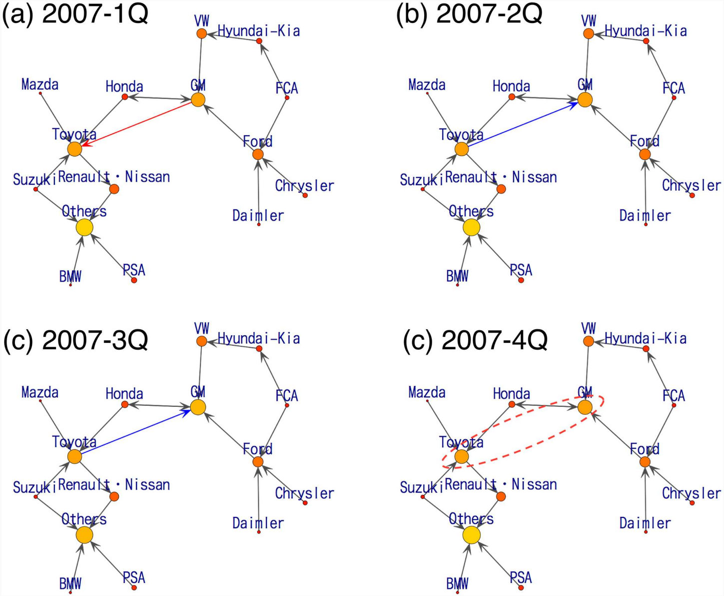
averages and standard deviations of sales shares



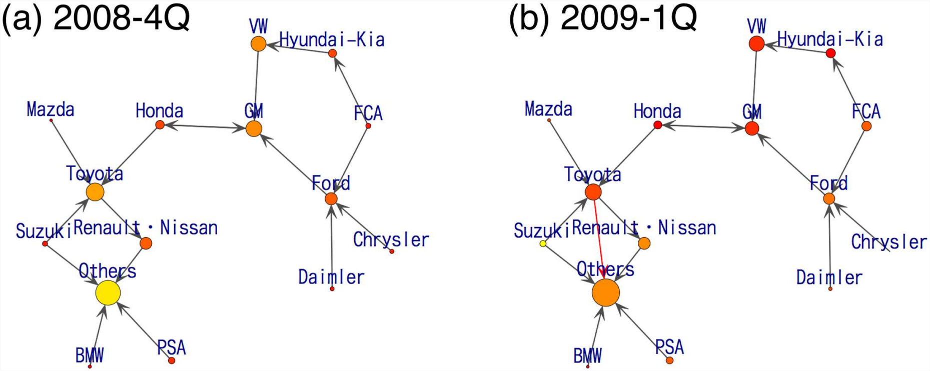
2007-1Q



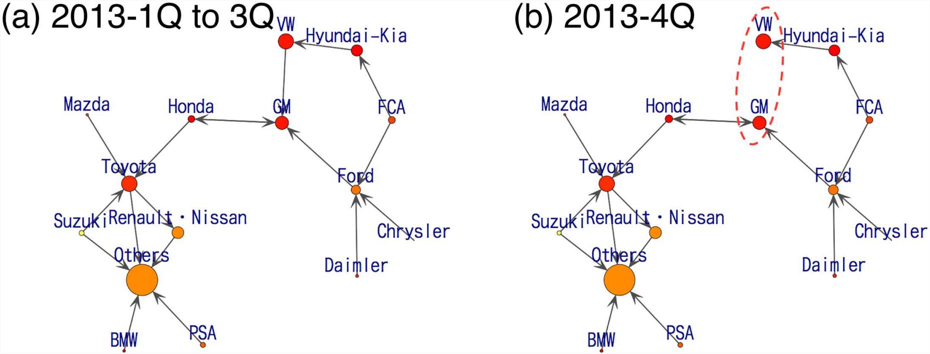
- remove minor edge below 0.24
- show market share with node size
- cf. GM and Honda are allied



In March 2008, TOYOTA has become the world's top seller by beating GM



In 2009, TOYOTA launched a massive recall



In 2013, VW beats GM in total sales amount to claim second position in the automobile industry

Conclusion

we presented the followings

- a model of transitions and stationary distributions
- a simple method for estimating transition matrices from a sequence of stationary distributions
- analysis of consumer transitions for sales share data without detailed recording of consumer transitions

further investigation would be devoted to

- other objectives and constraints to improve the accuracy of estimation and interpretability
- other probabilistic models for estimating changes in transitions

Chiba, Tomoaki et al. (Jan. 11, 2017). “Time-Varying Transition Probability Matrix Estimation and Its Application to Brand Share Analysis.” In: *PLoS ONE* 12.1, e0169981. DOI: 10.1371/journal.pone.0169981.