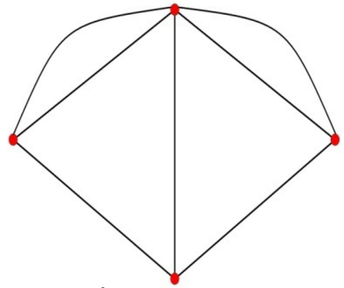
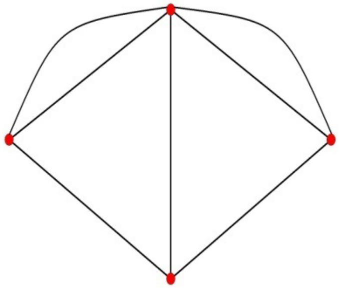


Modeling the Speed of Diffusion and Learning in Heterogeneous and Segregated Societies

SED July 3 2009   Plenary Talk of Matthew O Jackson

Based on Paper:
B. Golub and M.O. Jackson (2008)

Introduction: Speed



How quickly does information diffuse in a society?

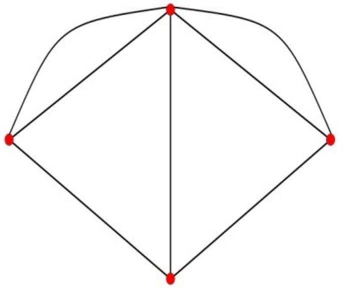
How quickly does a society reach an approximate consensus in opinions?

How quickly does a disease spread?

Financial contagions?

How do the answers depend on network structure?

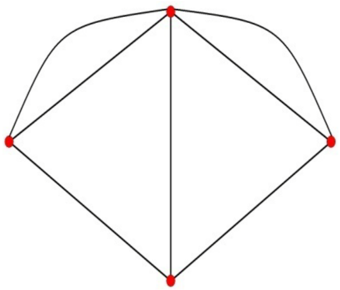
Main Questions



- How does speed depend on the process?
  - Pure diffusion – spreading out from one node
  - Updating/Learning – repeated processing of information from neighbors

- How does speed depend on the network?
  - Density of links
  - Pattern of links

Networks

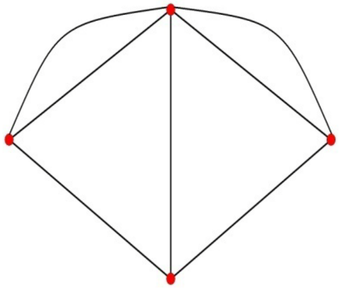


- Networks differ in their link density

- Networks differ in how links are spread across nodes: **Homophily**
  - Bias of relationships towards own type

- Technology and globalization are changing networks:
  - More relationships??
  - more/less homophily??

Outline

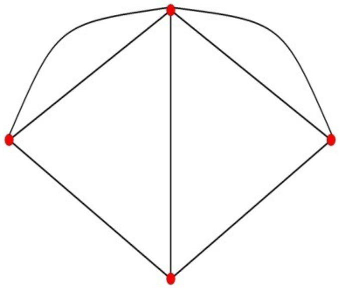


I Background on networks

II Modeling networks and diffusion/updating processes

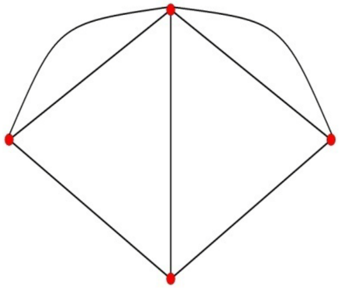
III Results on how structure impacts the speed of diffusion and learning

Density: Average Degree (# links)



HS Friendships (CJP 09)	6.5
Romances (BMS 03)	0.8
Borrowing (BDJ...)	3.2
Co-authors (Newman 01, GLM 06)	
Bio	15.5
Econ	1.7
Math	3.9
Physics	9.3
Facebook (Marlow 09)	120

Homophily/Segregation Patterns:



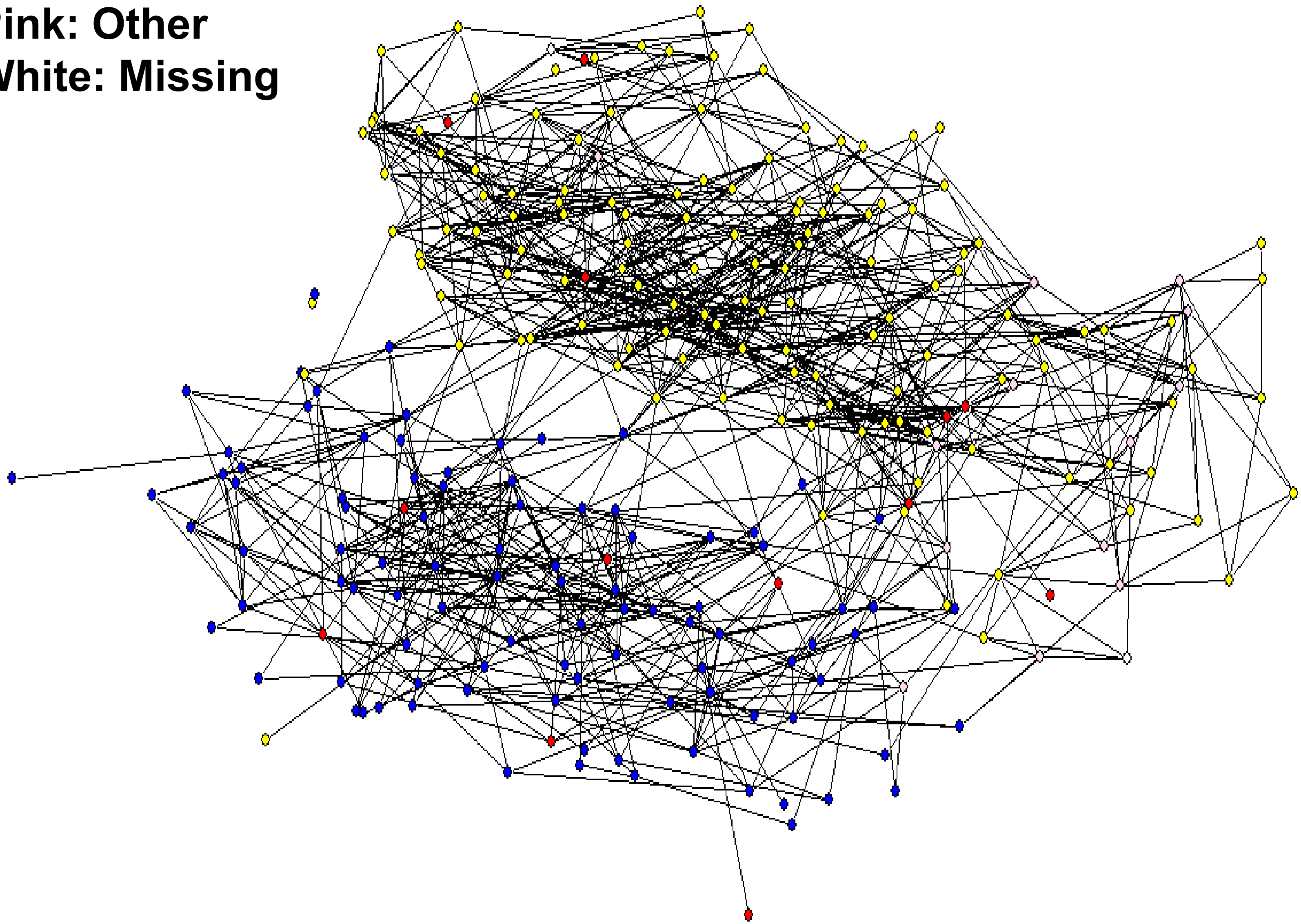
- Tendency to associate with others with similar characteristics: age, race, gender, religion, profession….

  – Lazarsfeld and Merton (1954) ``Homophily''

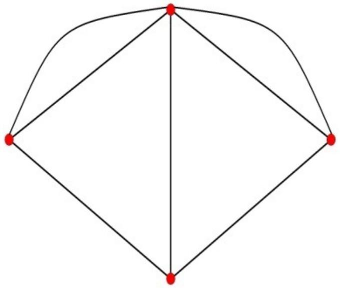
  – Shrum (gender, ethnic, 1988…), Blau (professional 1974, 1977), Burt, Marsden (variety, 1987, 1988), Moody (grade, racial, 2001…), McPherson (variety,1991…)…

Yellow: Whites
Blue: Blacks
Reds: Hispanics
Green: Asian
Pink: Other
White: Missing

Currarini, Jackson, Pin 2009

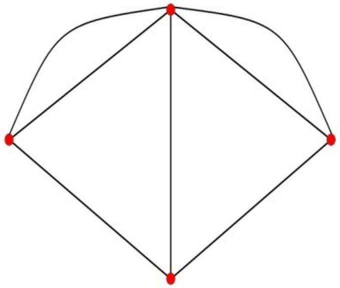


Adolescent Health, High School in US:



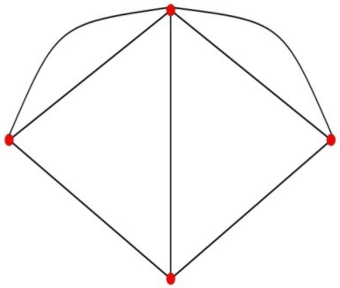
Percent:	52	38	5	5
	White	Black	Hispanic	Other
White	86	7	47	74
Black	4	85	46	13
Hispanic	4	6	2	4
Other	6	2	5	9
	100	100	100	100

How much more likely to link to own type?



84 schools	Race	Sex	Grade (Age)
Avg	1.6	1.5	8.5
stdev	.10	.08	.90

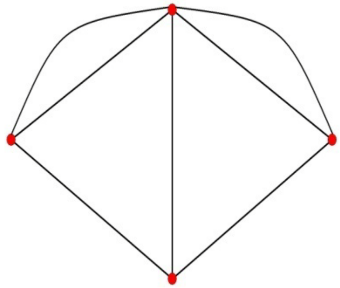
Modeling:



I Random Networks

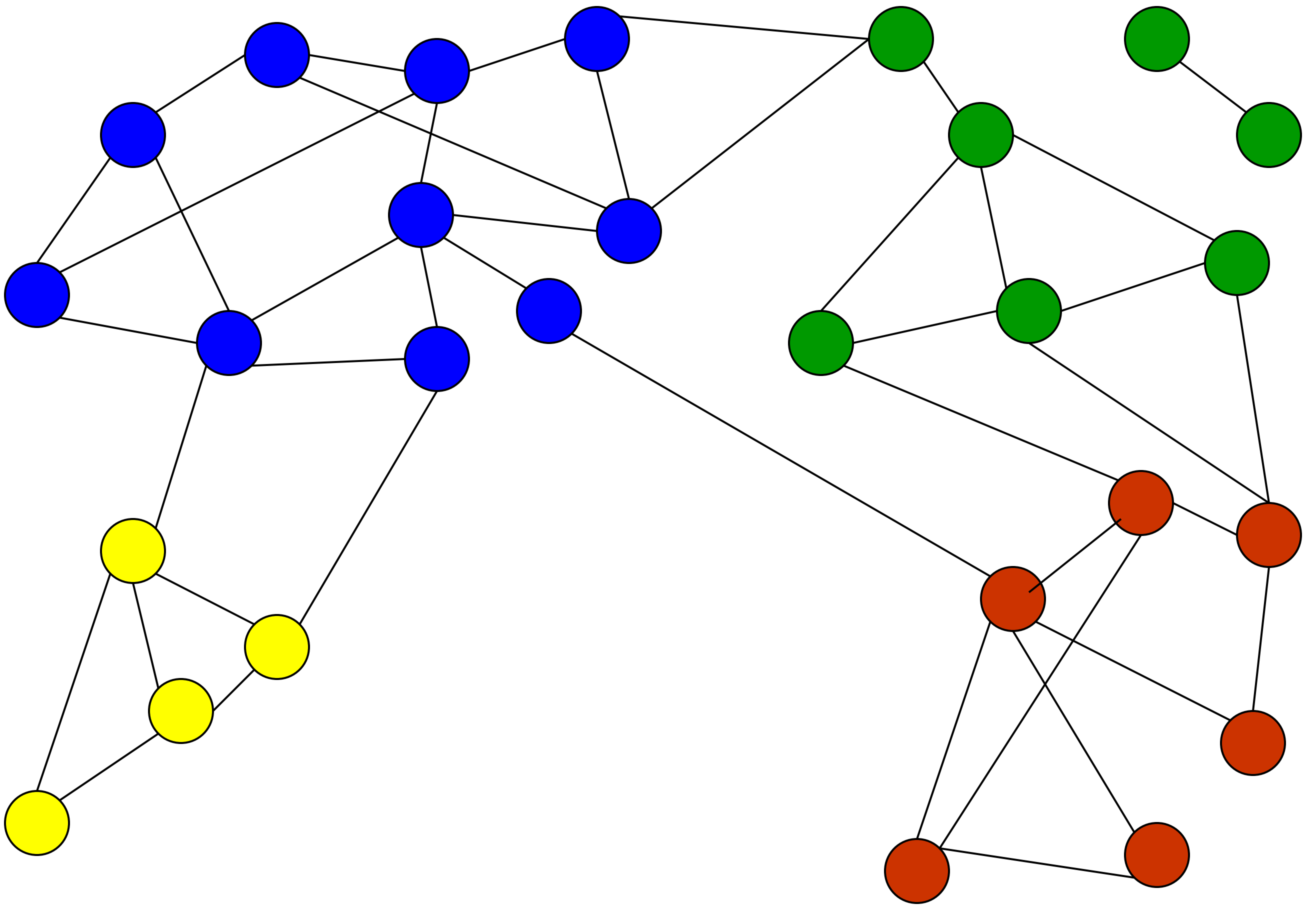
II Speed of Diffusion/Learning

Multi-Type Random Network Model

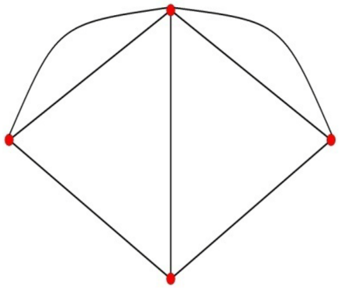


- $\{1, \ldots, n\}$ agents/nodes
- Partitioned into groups $N_1, \ldots, N_m$
- Node i in group k is linked to a node j in group k' with probability $P_{kk'}$ (undirected)
- Homophily: $P_{kk} > P_{kk'}$ for $k' \neq k$

Multi-Type Random network

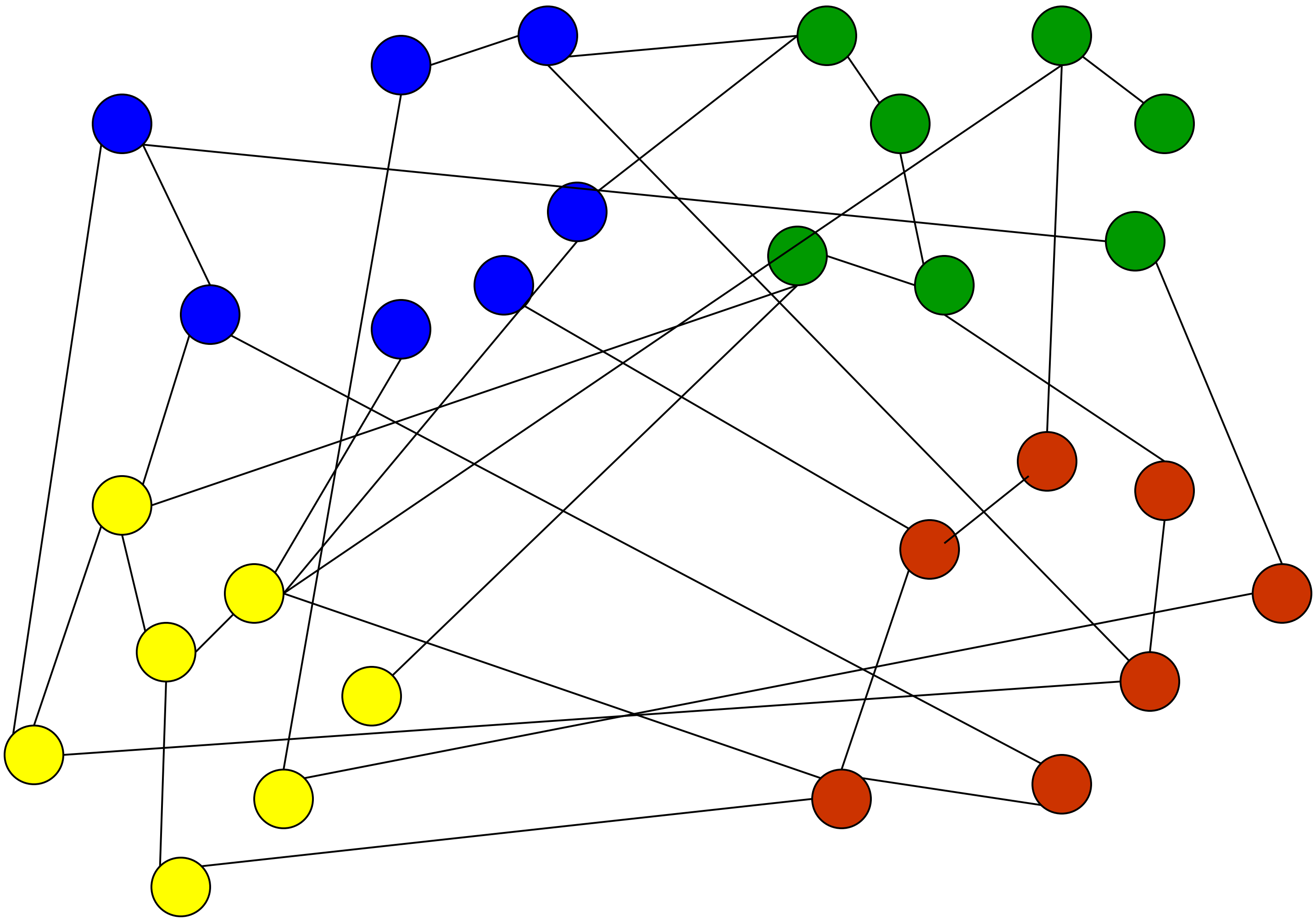


For Exposition: Islands Model

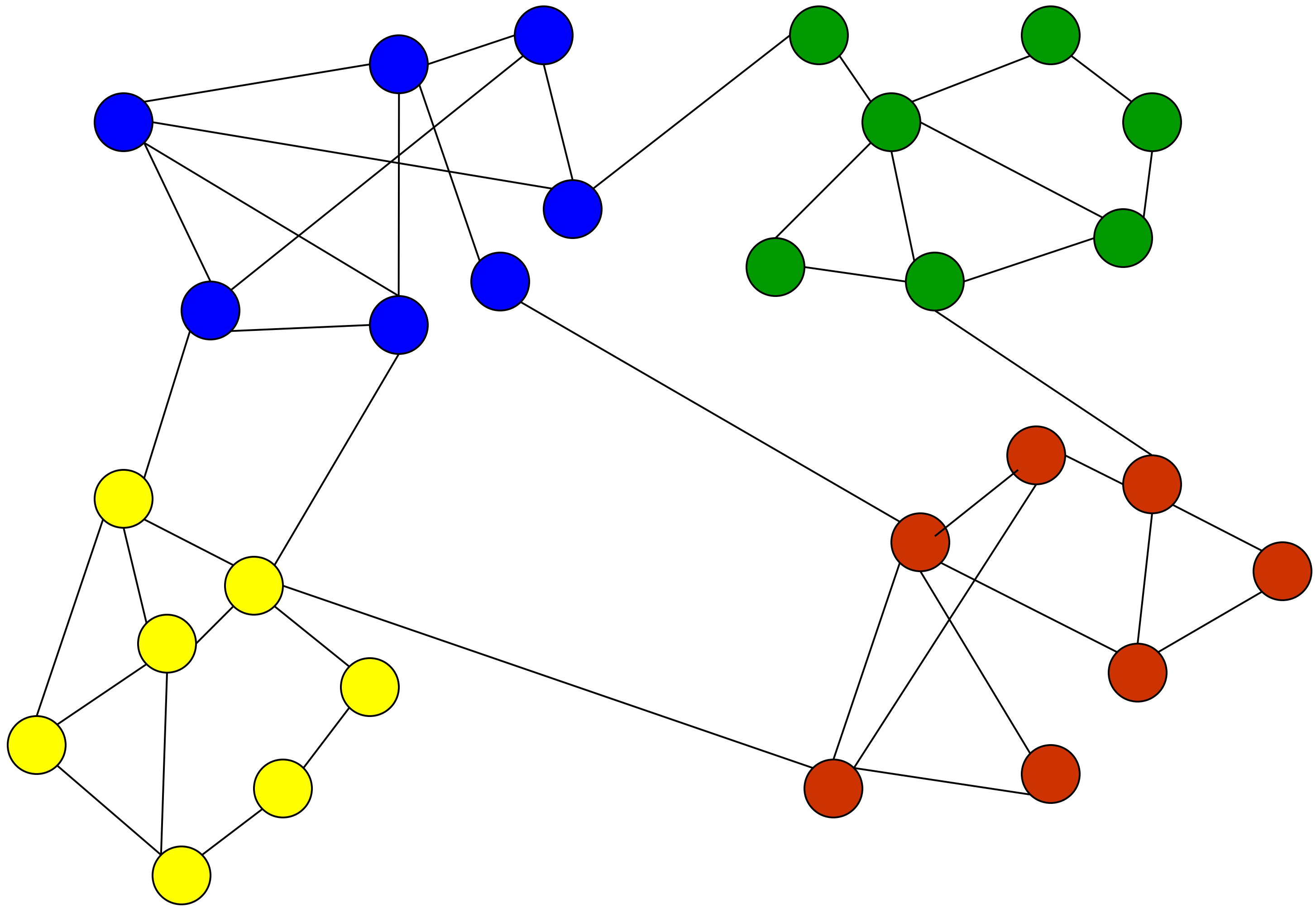


- Link with a same type agent: probability p_s
- Link with a diff type agent: probability p_d
- Overall probability of links p (>> 1/n)
- m equal-sized groups

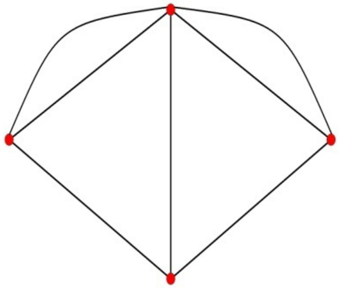
Example Low Homophily



Example High Homophily



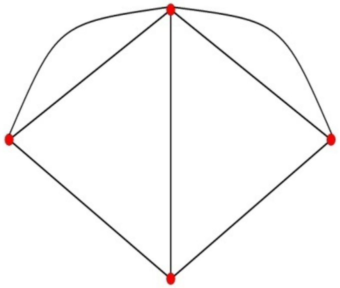
Diffusion/Learning Processes:



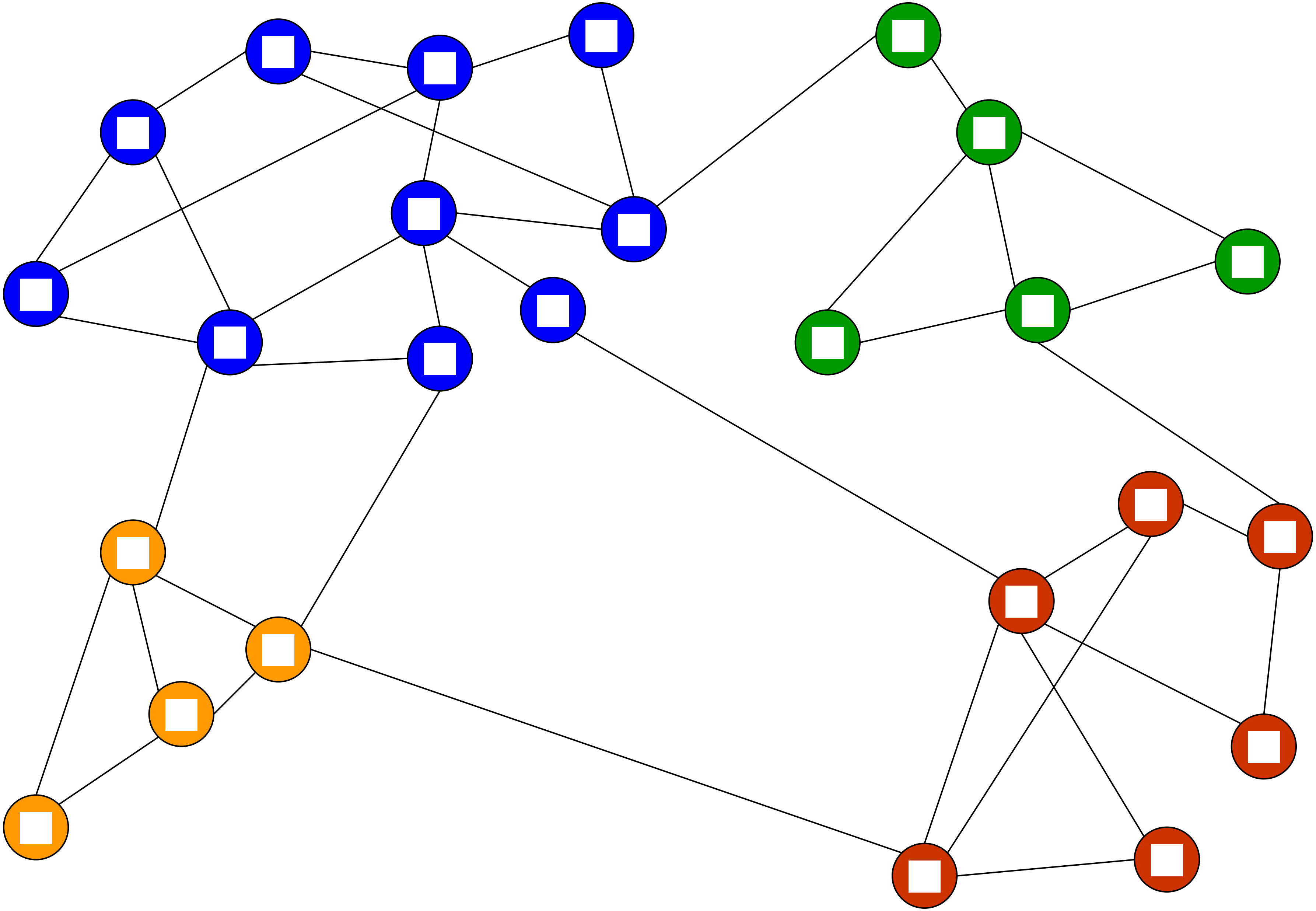
(1) Diffusion / gossip / broadcasting / navigation: speed depends on shortest path structure

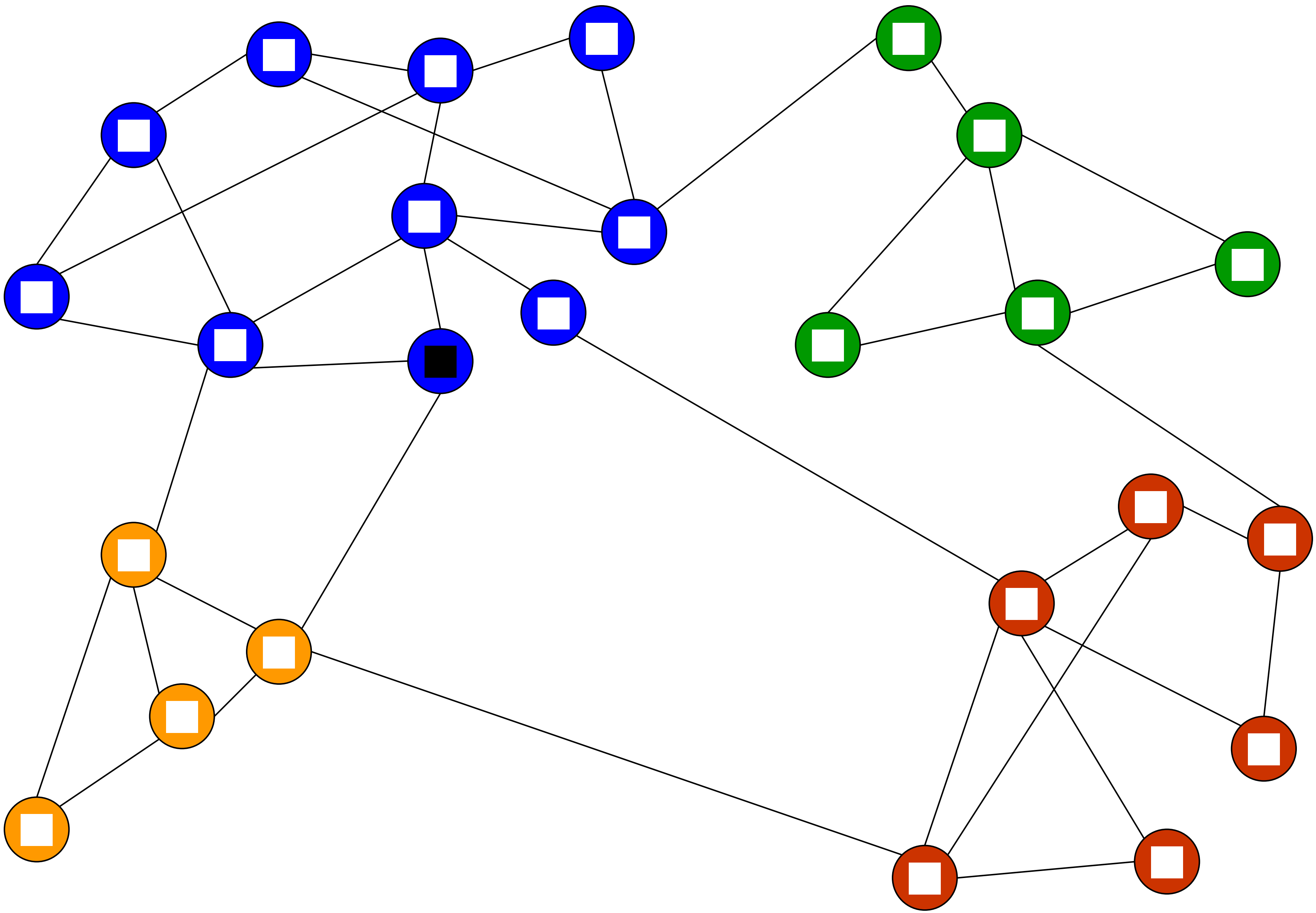
(2) Updating: repeated discussion and weighing of neighbors' opinions, iterative best response...

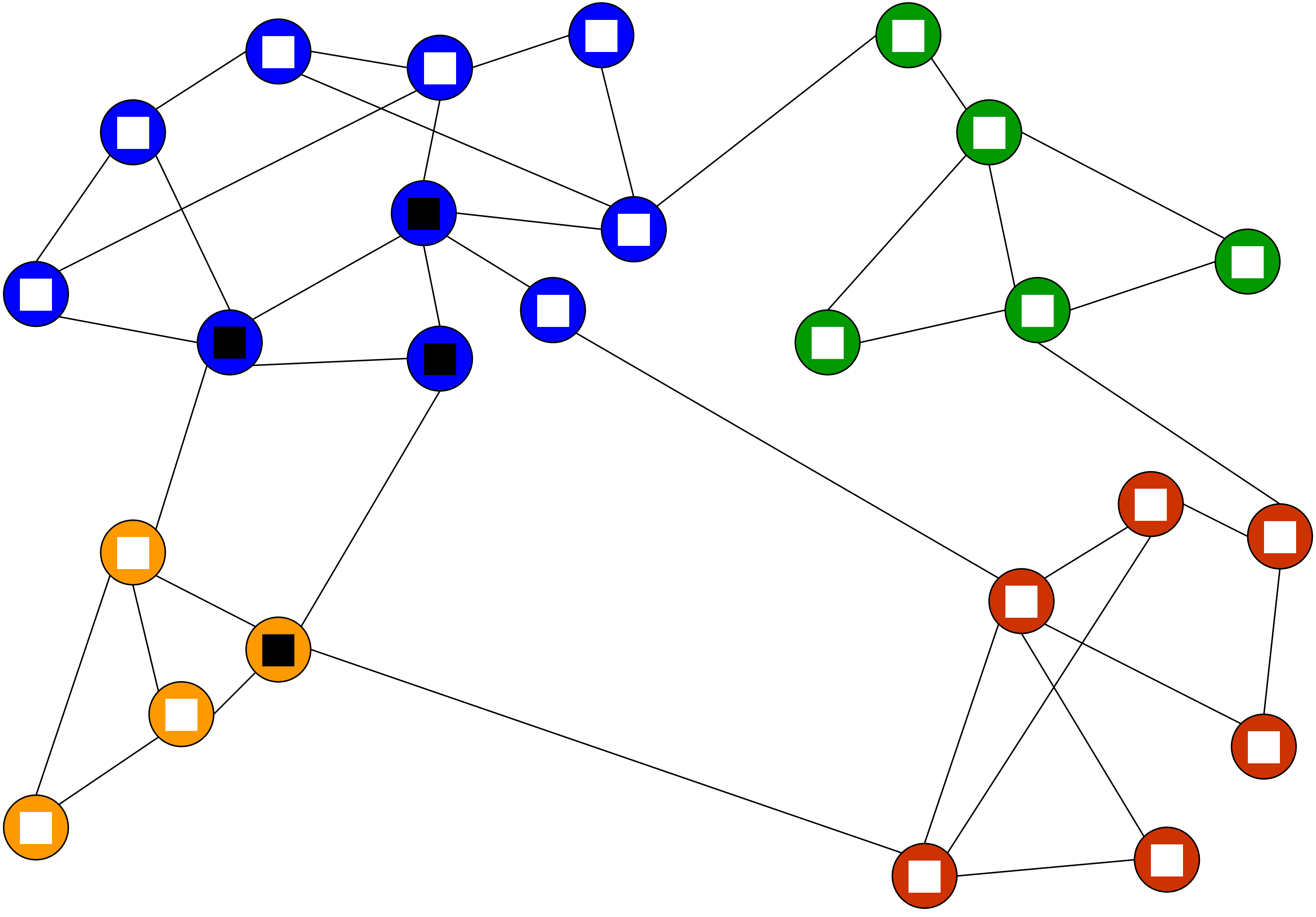
1. Shortest Paths / Pure Diffusion

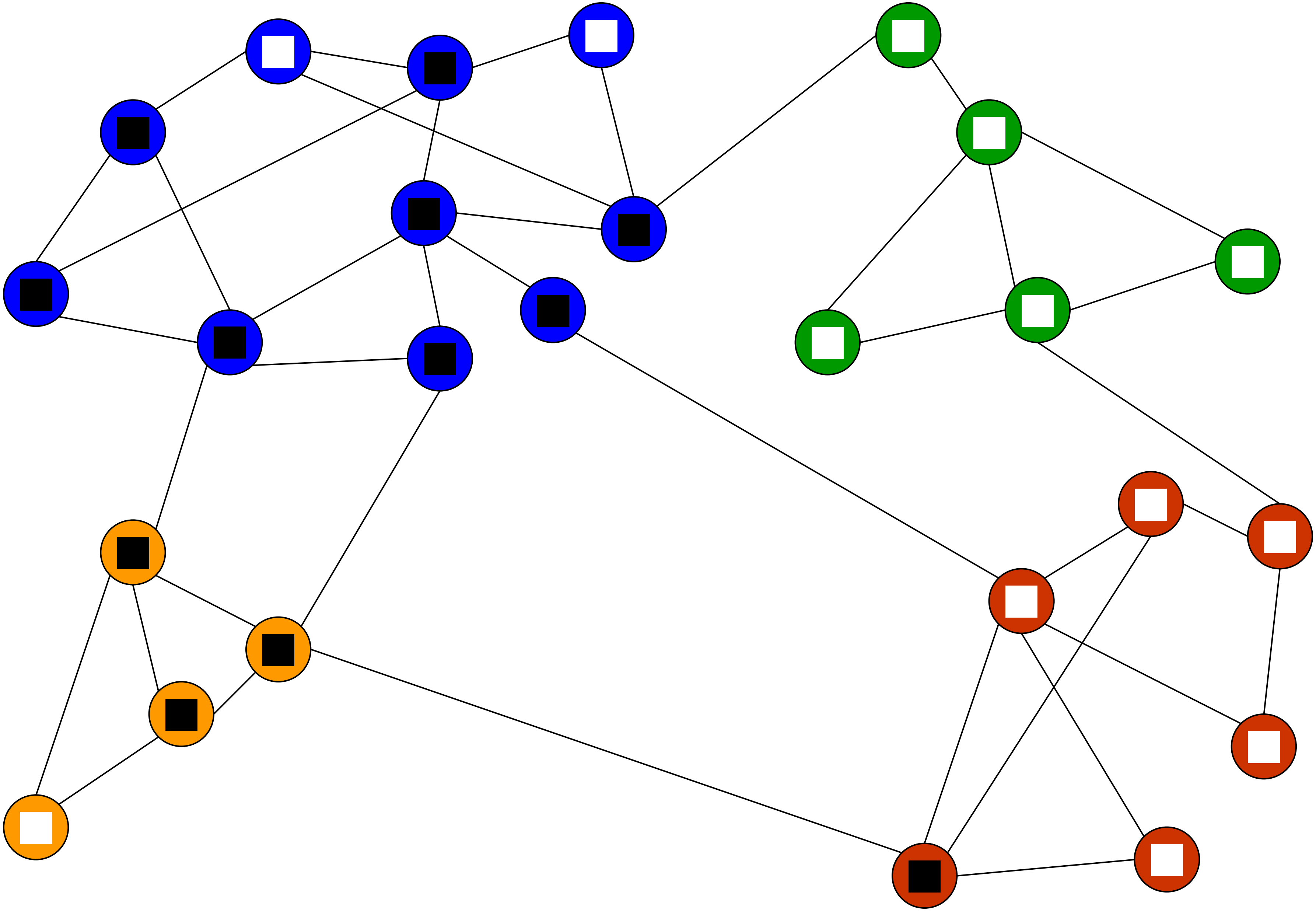


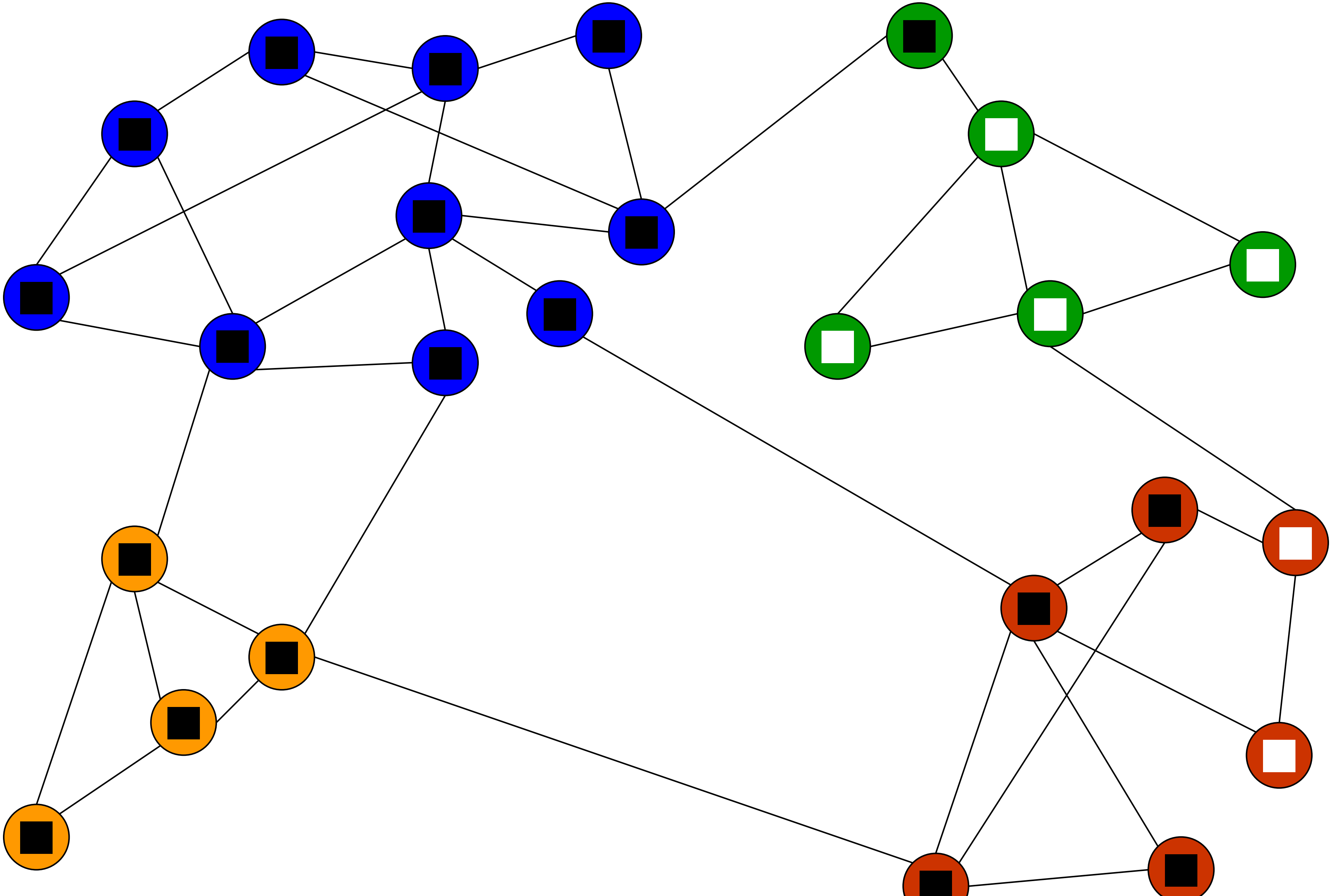
- Characterize how shortest paths are affected by density and homophily

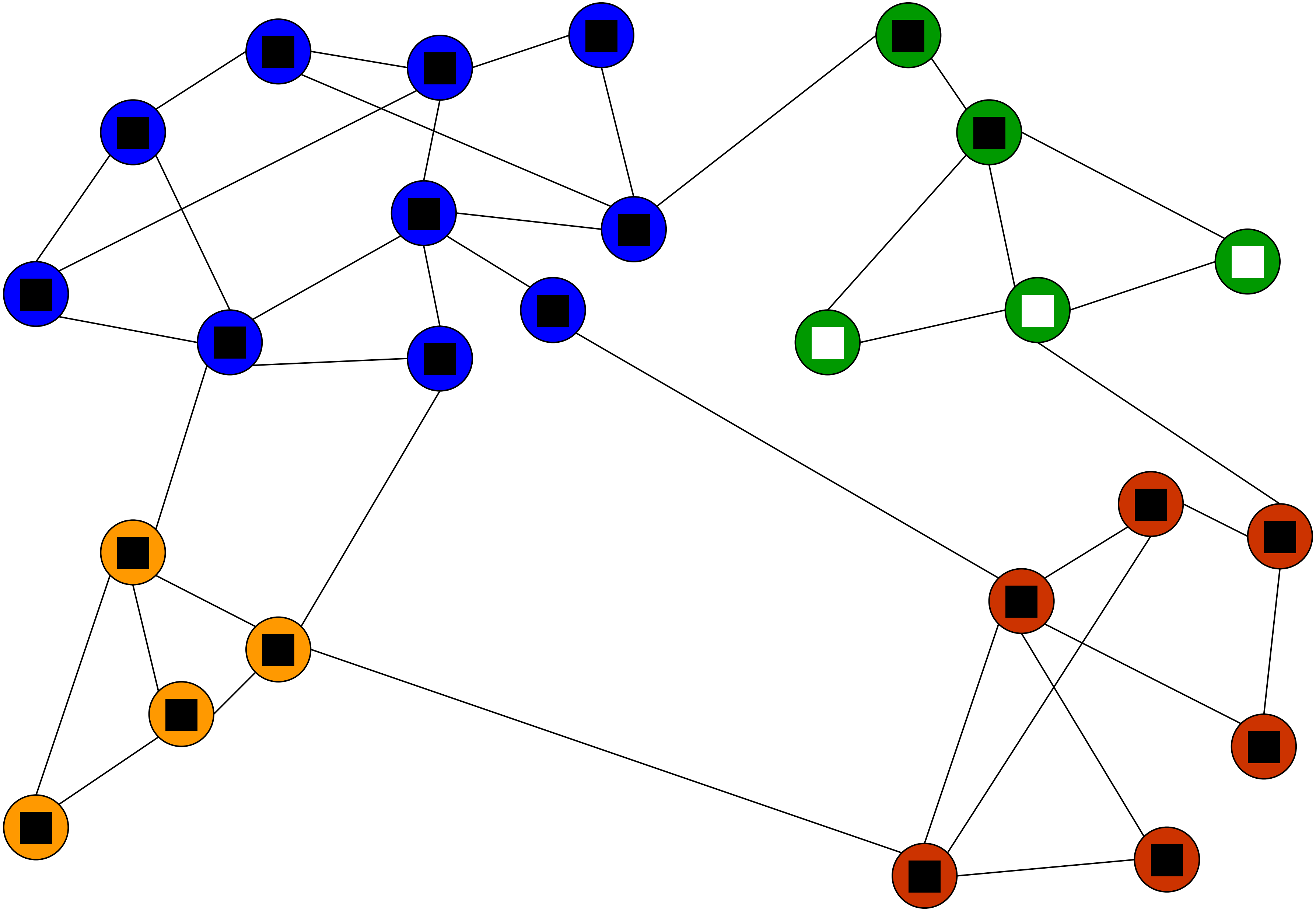


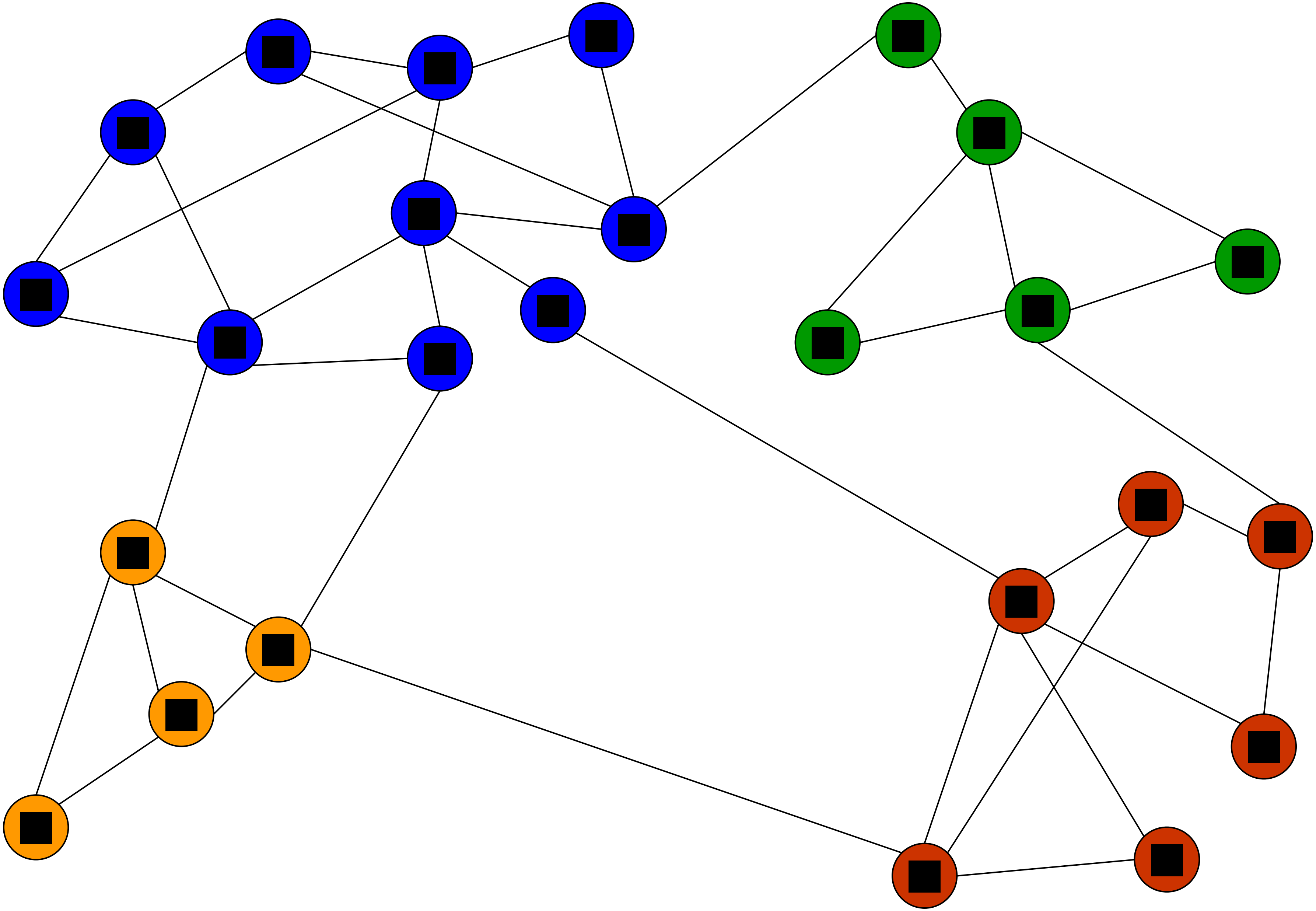








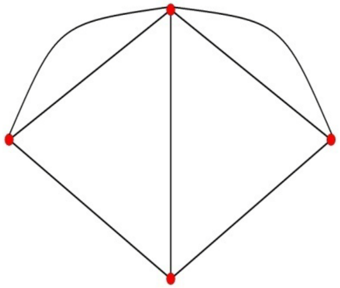




2. Model of Updating/Learning
French (1956), Harary (1959), DeGroot (1974)...

- At each date talk to neighbors
- update opinion or behavior by taking an average of neighbors' opinion/behavior
- Iterate on this process

DeGroot Model



- Individuals {1, ... n} are in a network

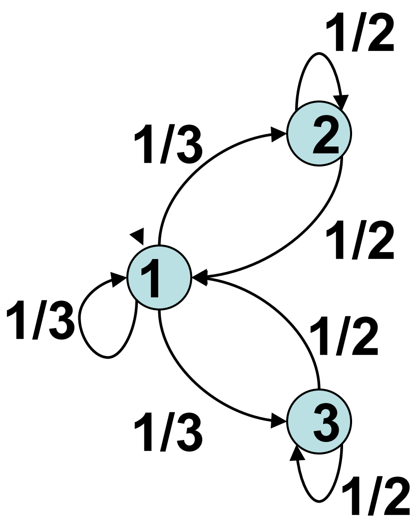
- $T_{ij} = 1/d_i$ if i is linked to j and = 0 otherwise

- Start with beliefs (behavior, etc.) $b_i(0)$ in [0,1]

- Updating: $b_i(t) = \sum_j T_{ij}\, b_j(t-1)$

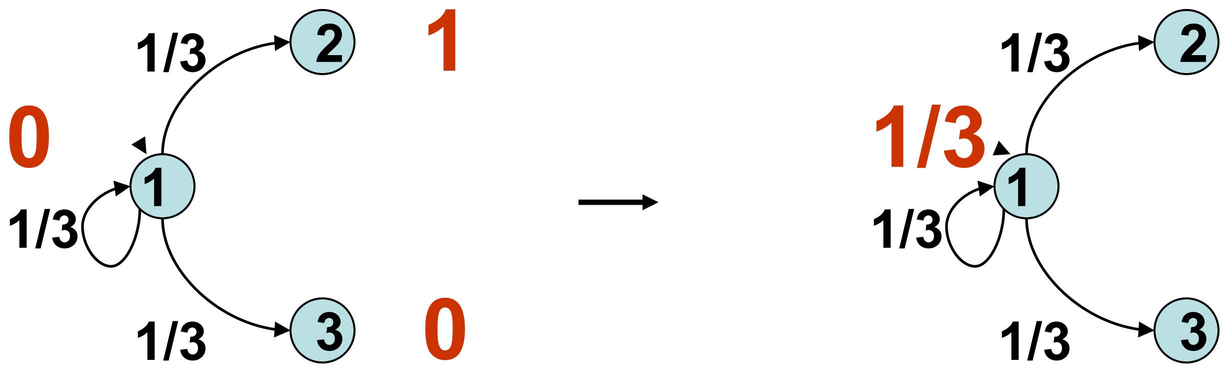
  So: $b(t) = \mathbf{T}\, b(t-1) = \mathbf{T}^t\, b(0)$

Example

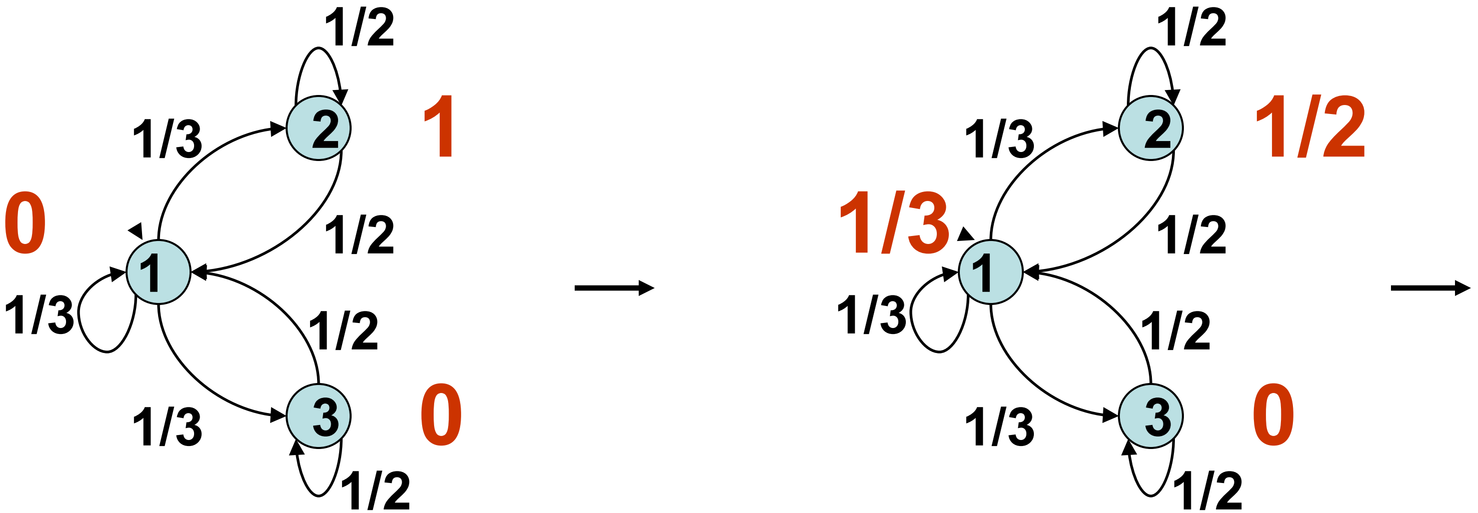


$$T = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

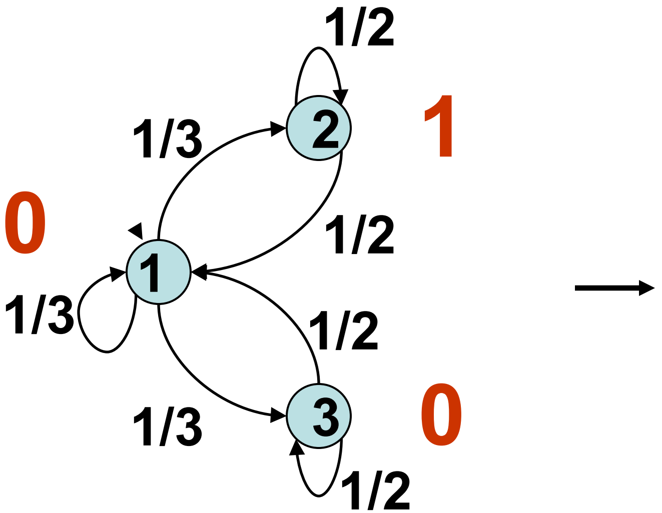
Updating

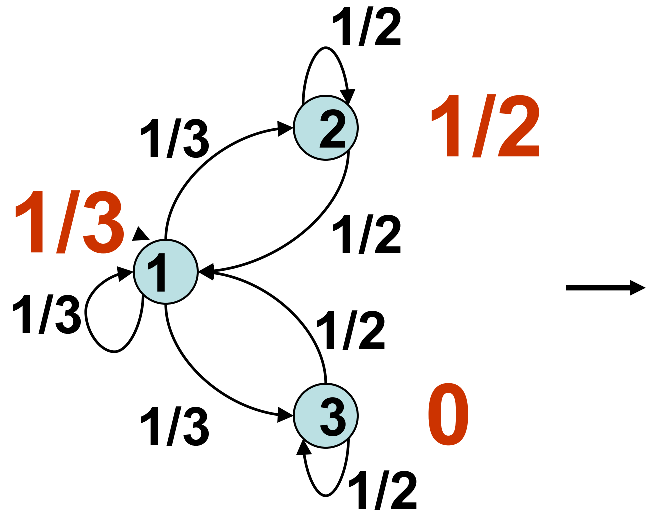


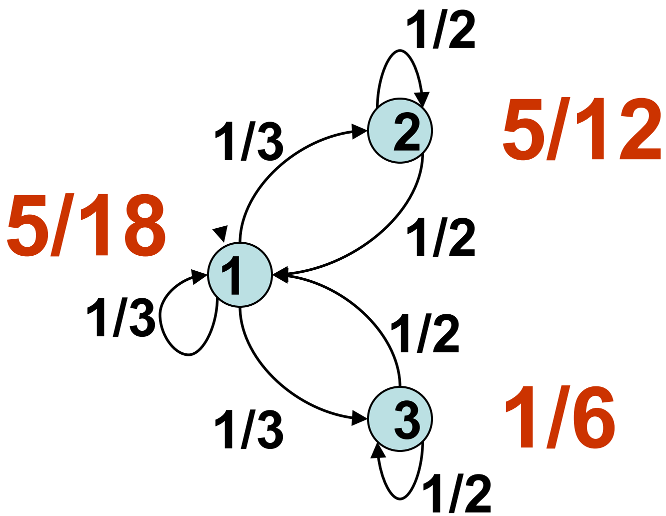
Updating



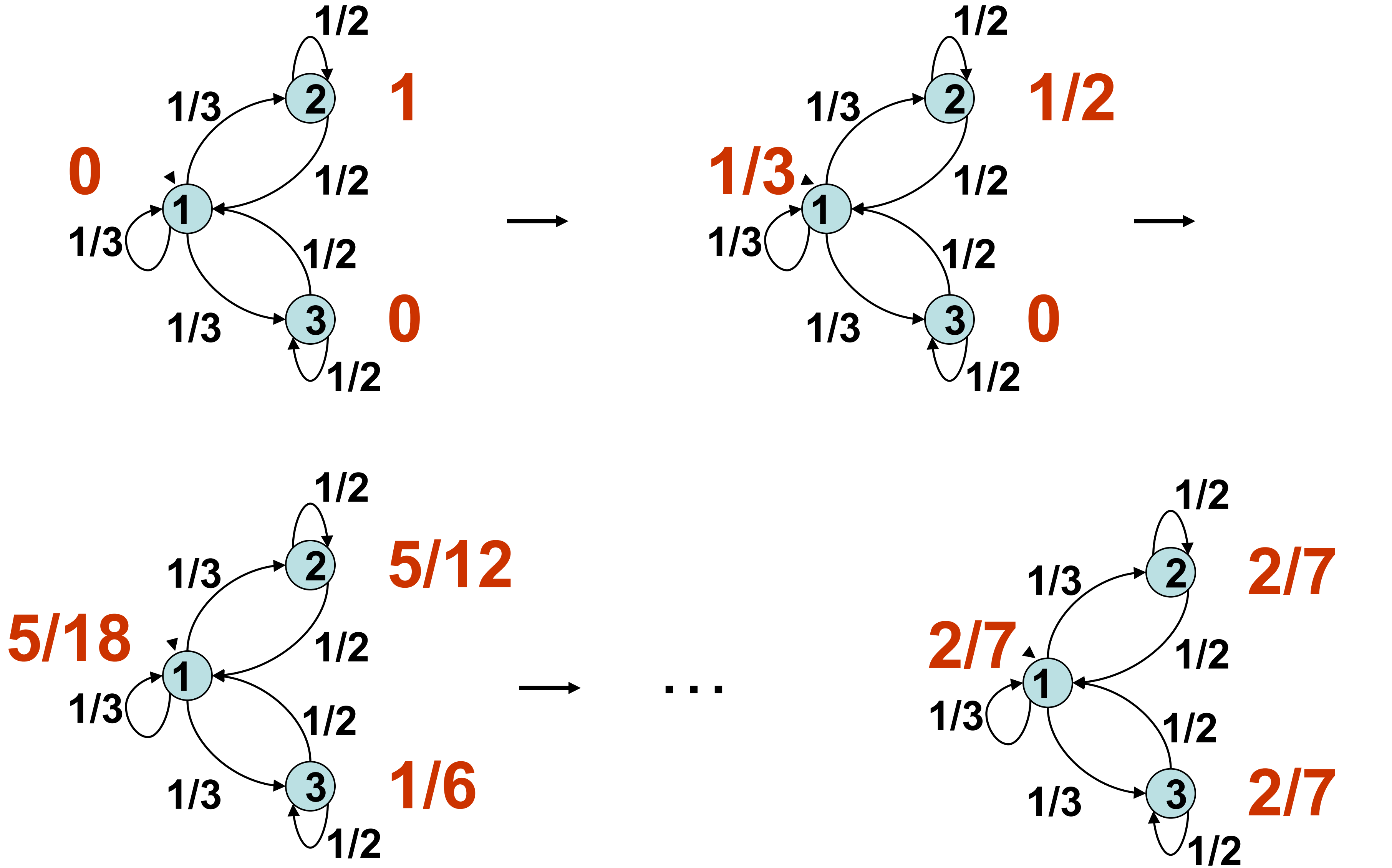
Updating


1/2
1/3
2
1
0
1/2
1
1/3
1/2
1/3
3
0
1/2


1/2
1/3
2
1/2
1/3
1/2
1
1/3
1/2
1/3
3
0
1/2

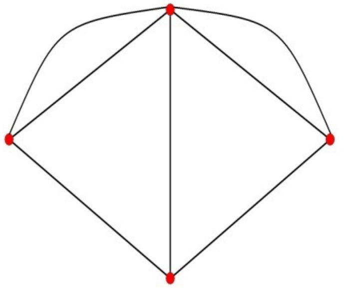

1/2
1/3
2
5/12
5/18
1/2
1
1/3
1/2
1/3
3
1/6
1/2

Updating



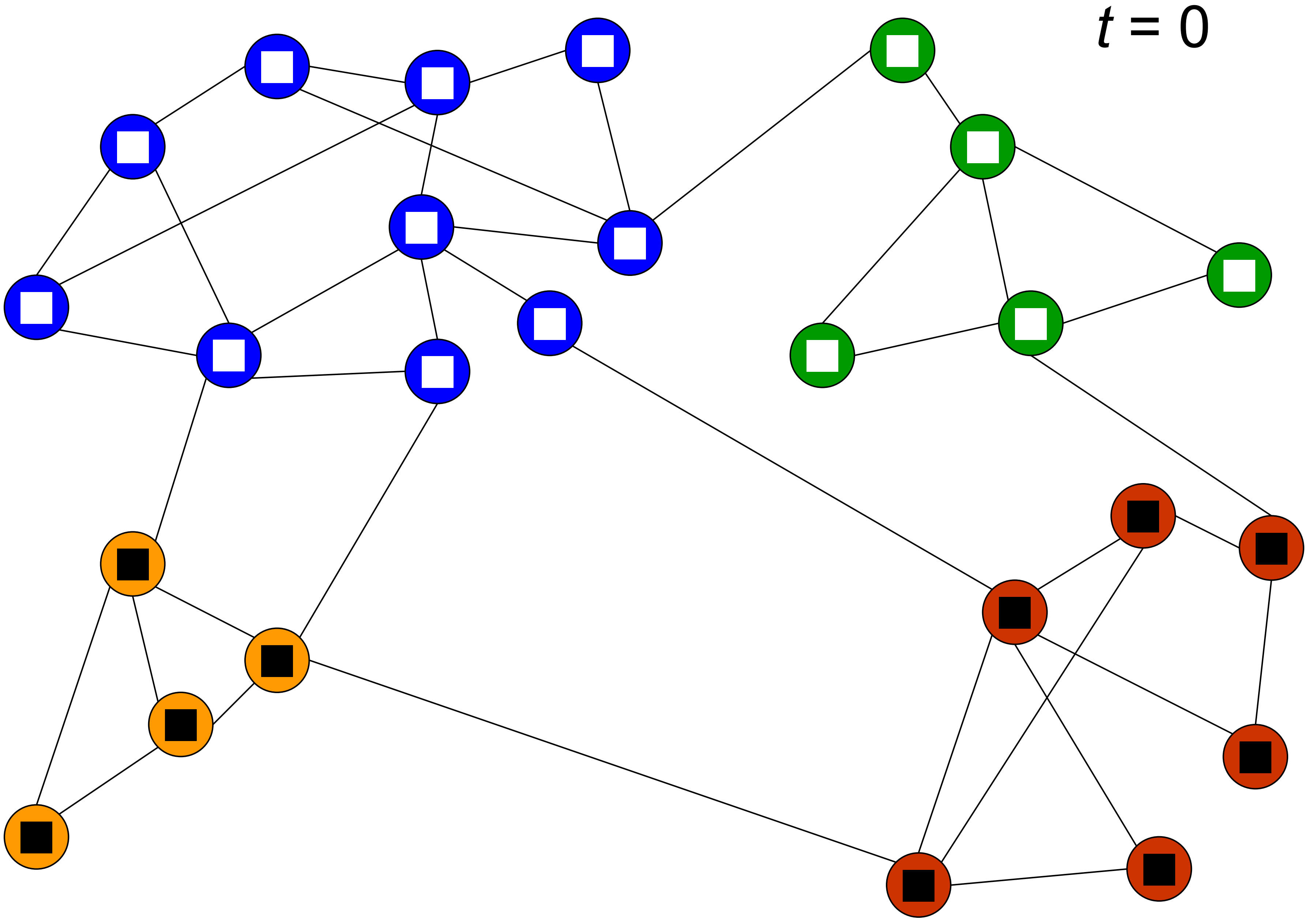
1/2
1/3
2
1
0
1/2
1
1/3
1/2
1/3
3
0
1/2
1/2
1/3
2
1/2
1/3
1/2
1
1/3
1/2
1/3
3
0
1/2
1/2
1/3
2
5/12
5/18
1/2
1
1/3
1/2
1/3
3
1/6
1/2
. . .
1/2
1/3
2
2/7
2/7
1/2
1
1/3
1/2
1/3
3
2/7
1/2

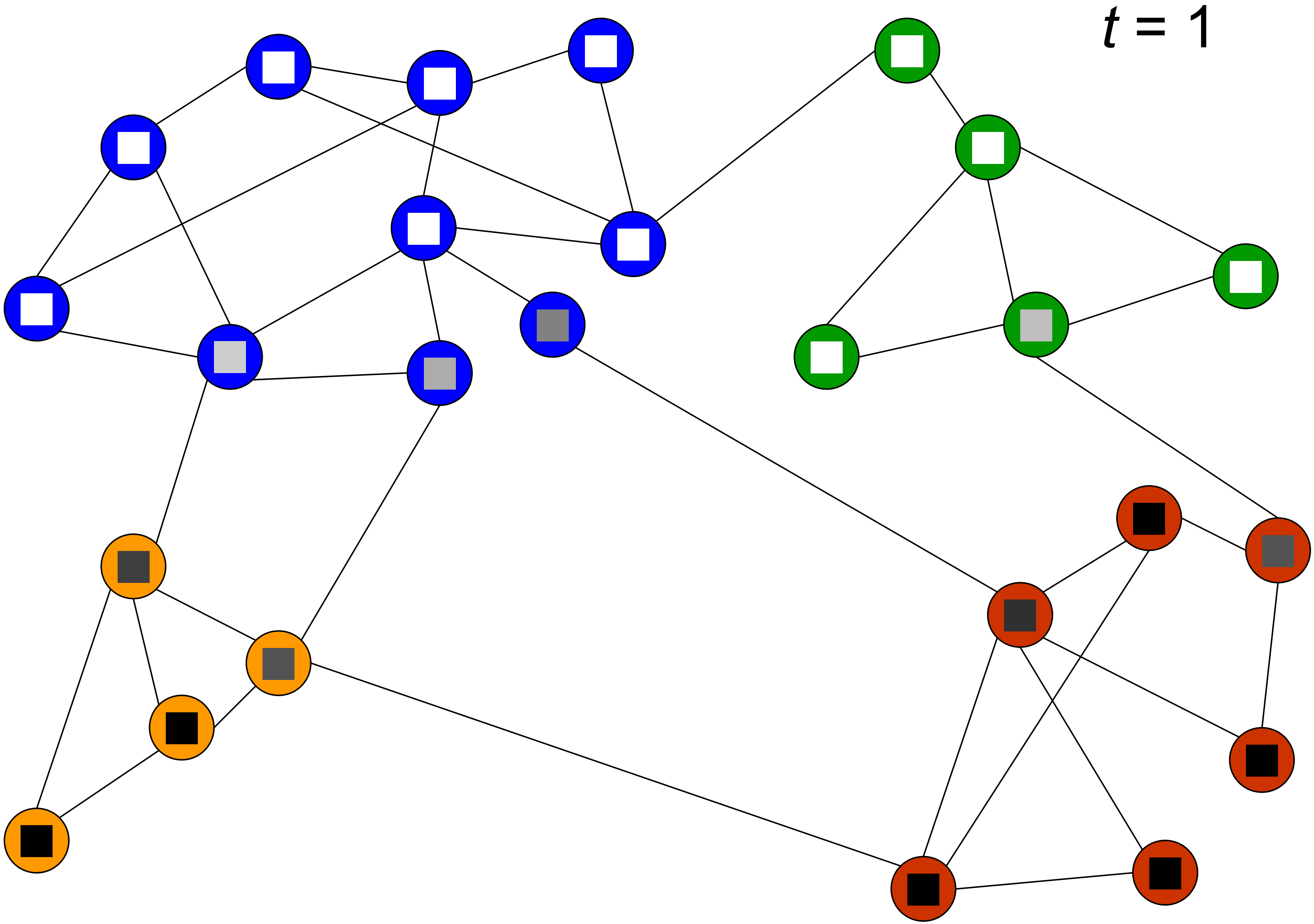
Convergence

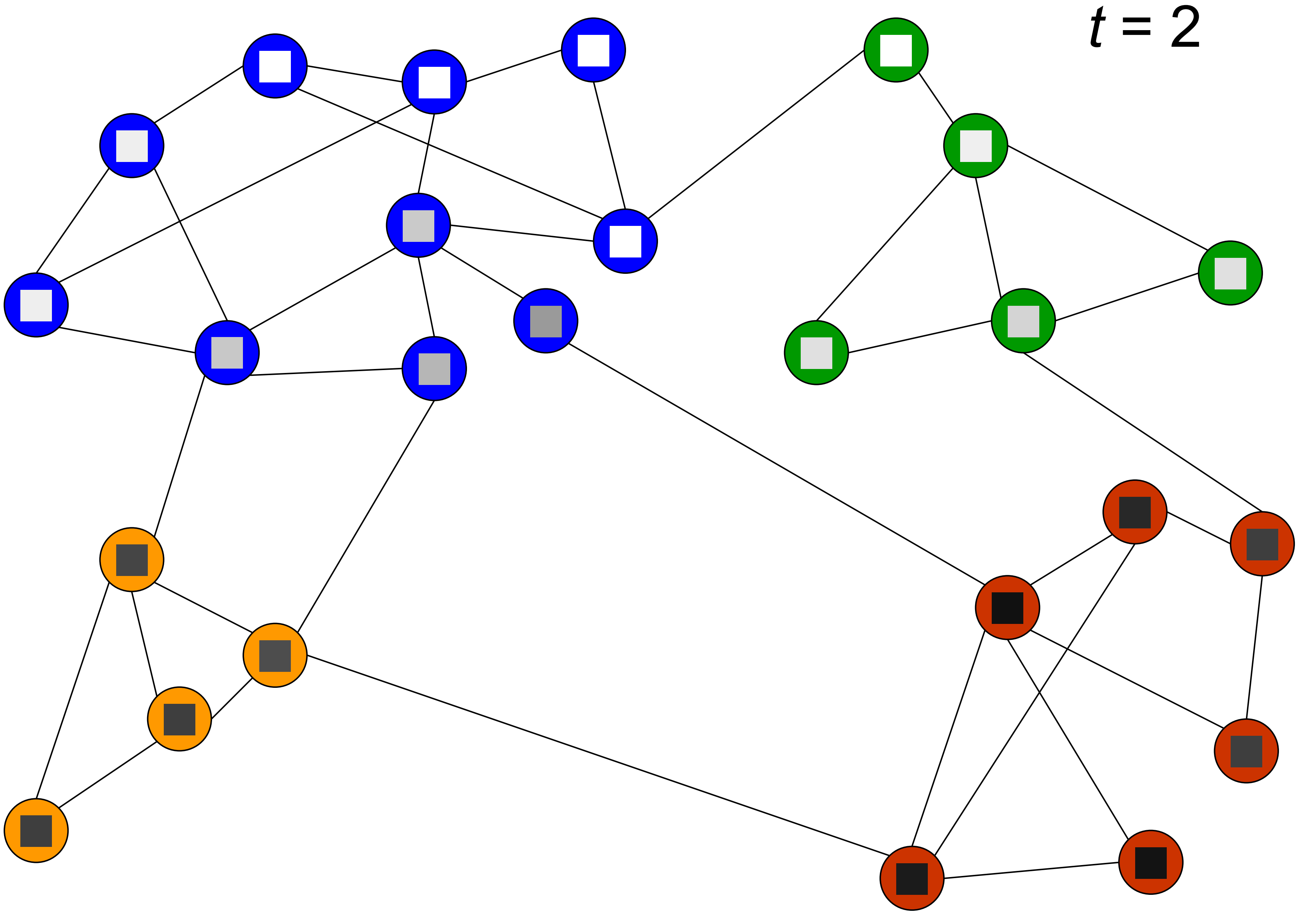


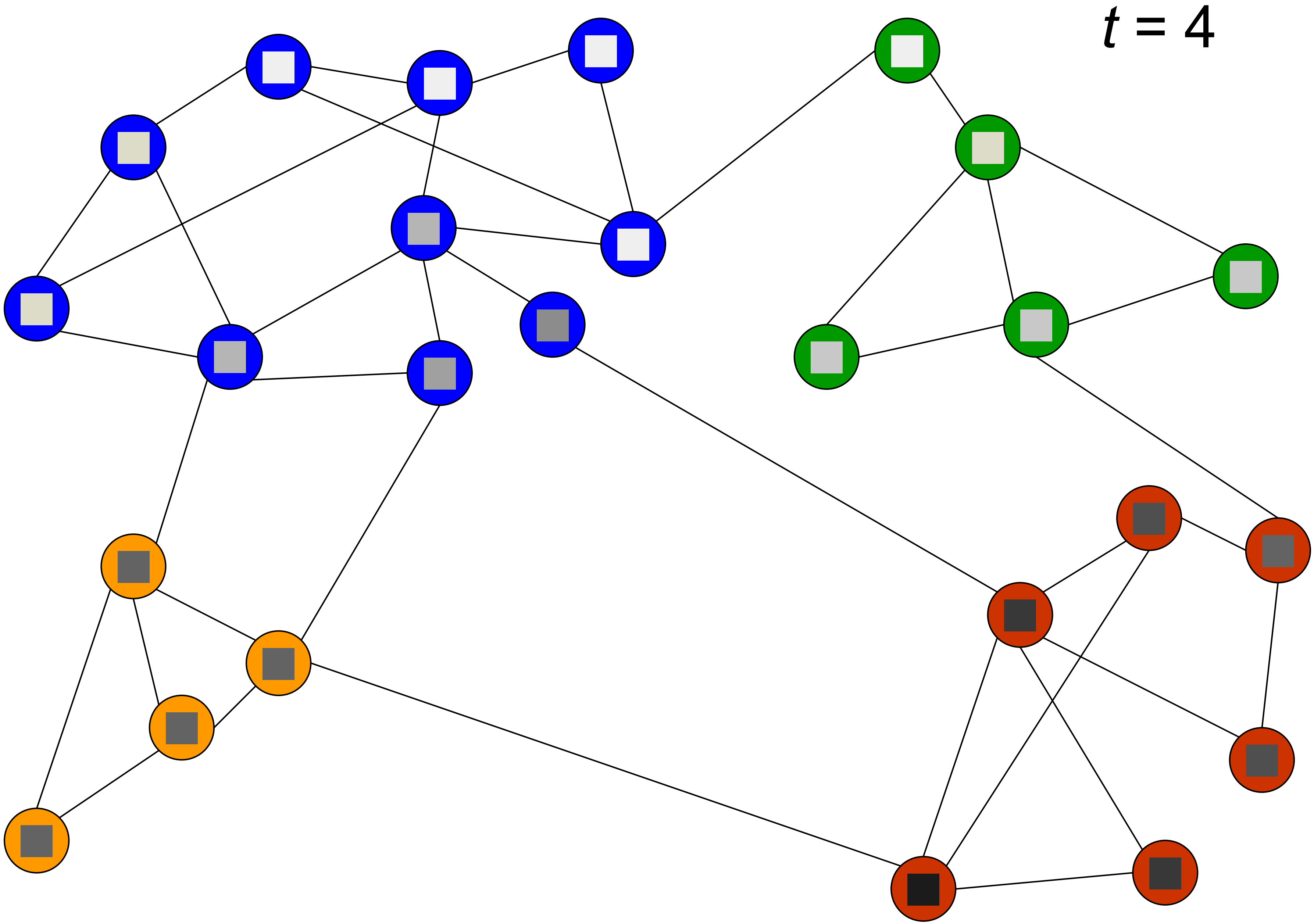
If the network is path-connected and and at least one agent places some weight on own opinion, then the DeGroot process has a unique limit (a standard Markov result) and all agents converge to same belief

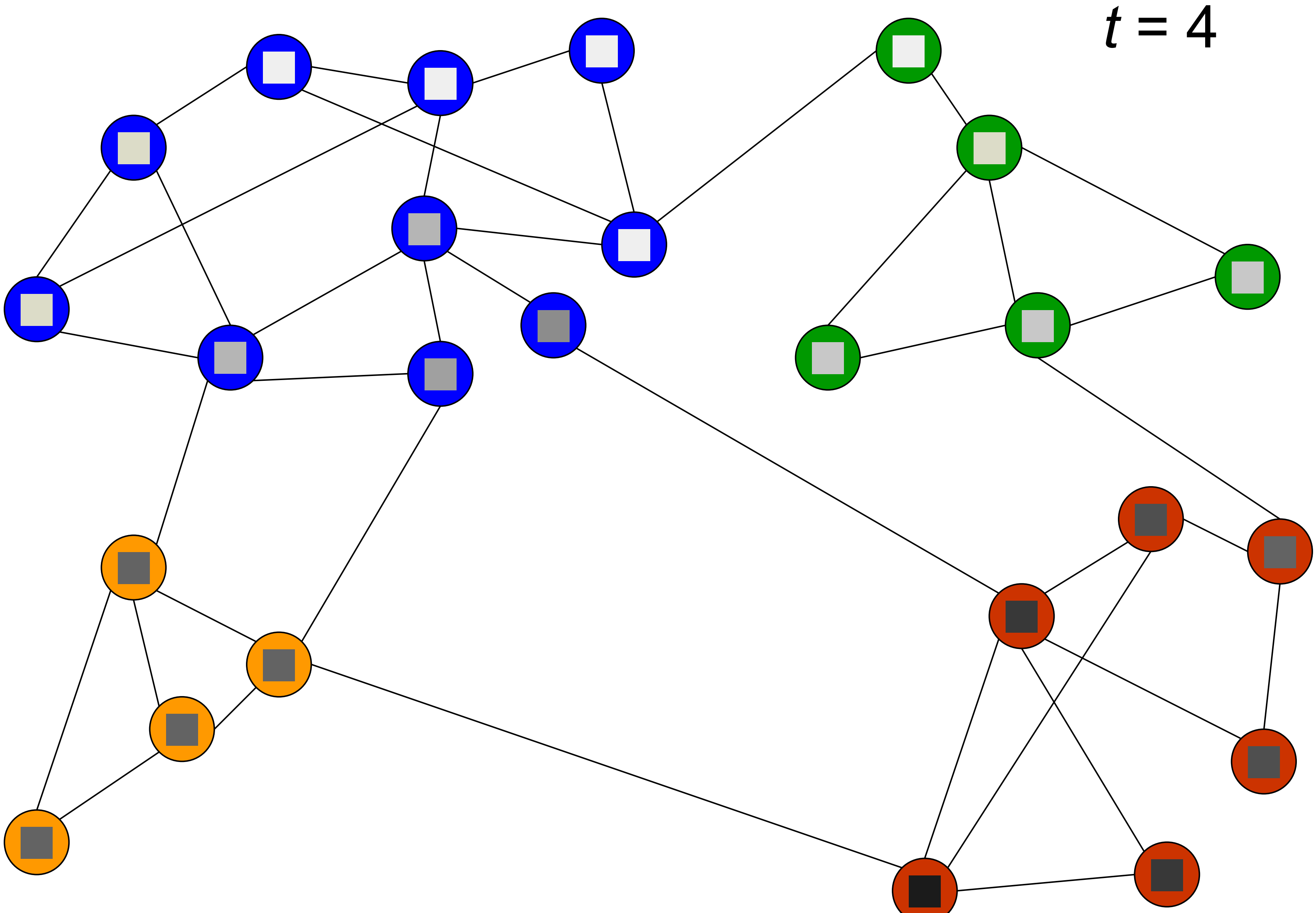
The Consensus is $\sum_i \mathbf{b_i(0)\ d_i / \sum_j d_j}$


$t = 0$

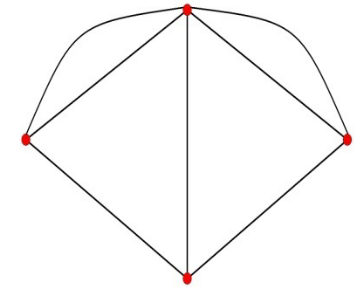

$t = 1$


$t = 2$


$t = 4$


$t = 4$

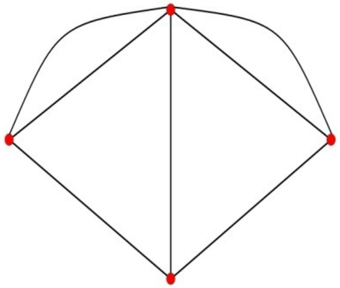
Other Processes that are equivalent to DeGroot:



- **Myopic best responses**: Each agent wants to match the average behavior of his/her neighbors –

- **Random walk – Markov Chain**: Start at some node/state and randomly transition - how long until distribution of location looks like steady-state distribution?

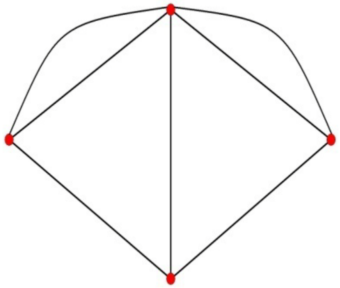
Consensus Time



$$CT(T,\varepsilon) = \sup_b \min \{t: ||T^t b - T^\infty b|| < \varepsilon\}$$

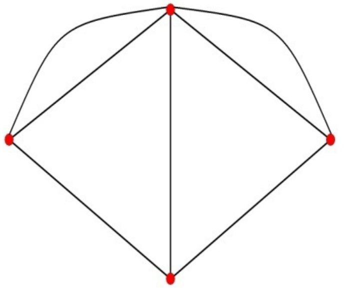
How long until vector of beliefs is within ε of its limit? (worst case)

Analysis



- Multi-Type Random Network is Drawn
- Diffusion: What is average path length?
- Learning/Random Walk: How fast does DeGroot process converge?

Theorem on Network Structure (Jackson 08)

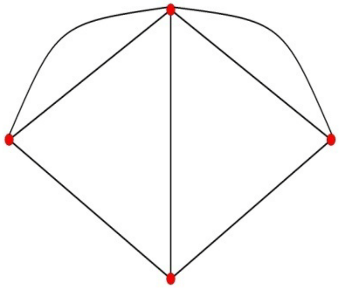


Large n results, let $d = np$ (avg degree)

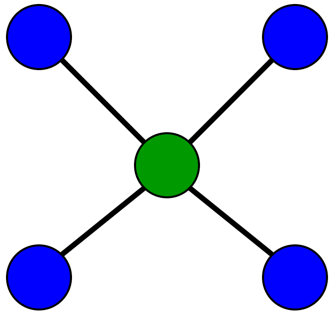
$$\frac{\text{AvgDist}(n)}{\log(n)/\log(d)} \to^{P} 1$$

link density matters but not homophily

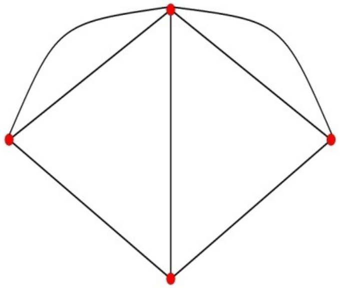
Intuition:

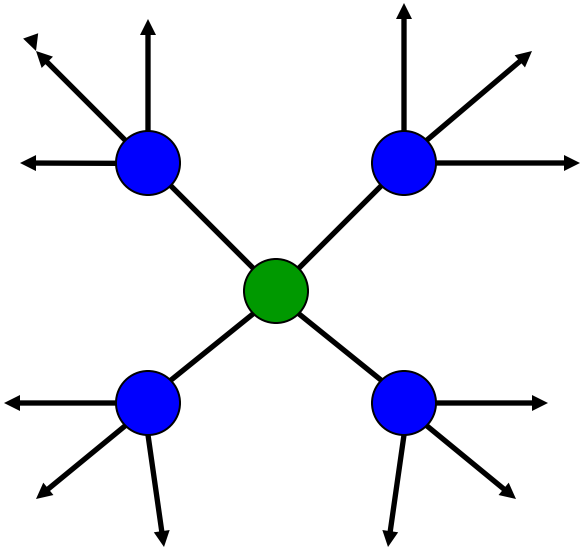


1 step: Reach d nodes,



Ideas:

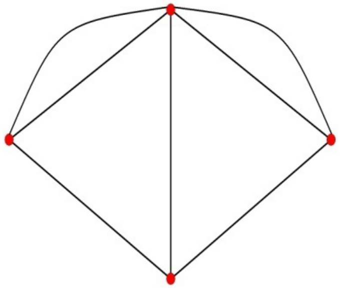


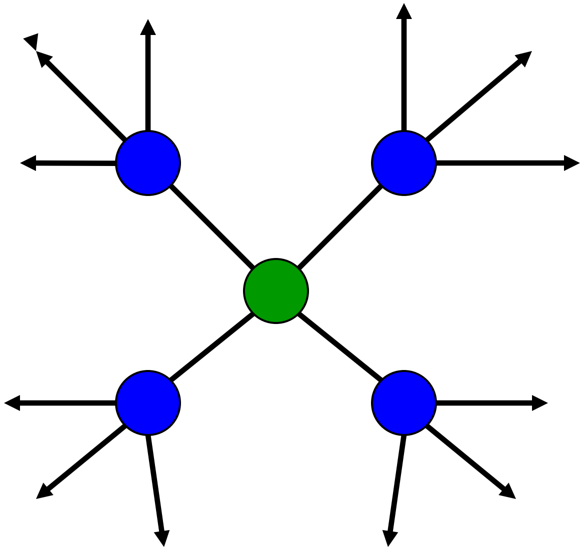


1 step: Reach d nodes,

then d(d-1),

Ideas:



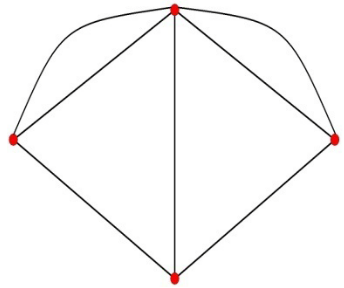


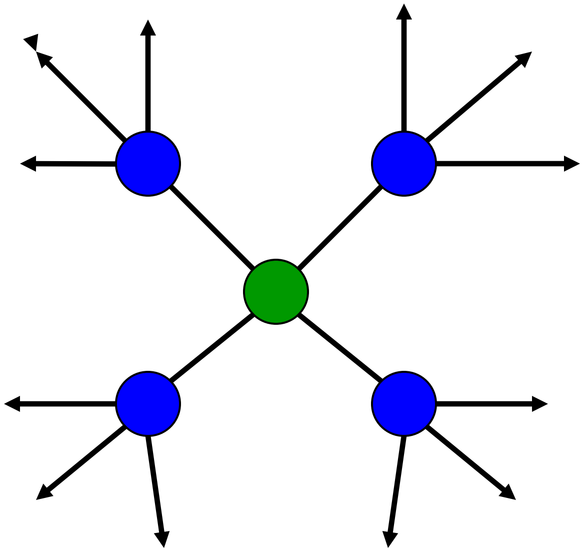
1 step: Reach d nodes,

then d(d-1),

then $d(d-1)^2$,

Ideas:





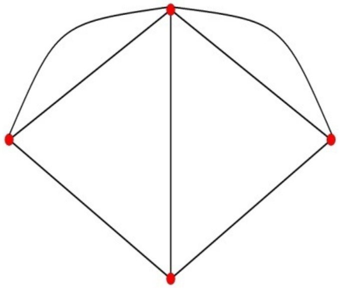
1 step: Reach d nodes,

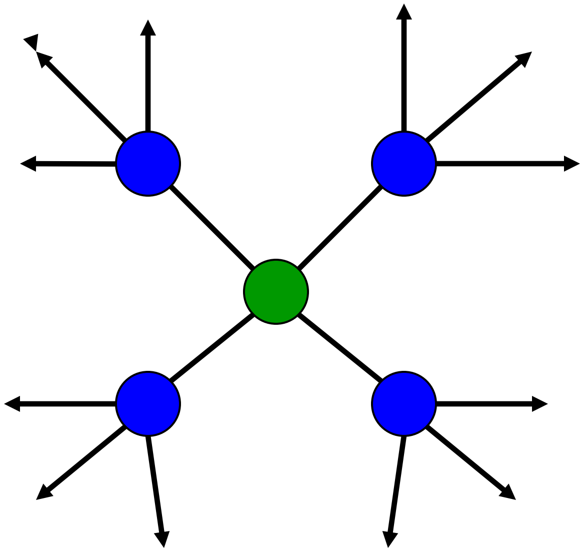
then d(d-1),

then $d(d-1)^2$, $d(d-1)^3$, ...

After k steps, totals roughly d^k

Ideas:



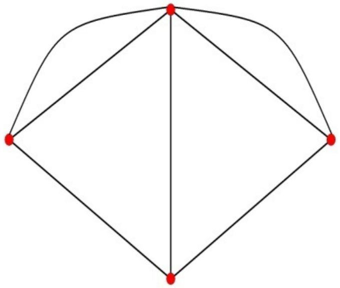


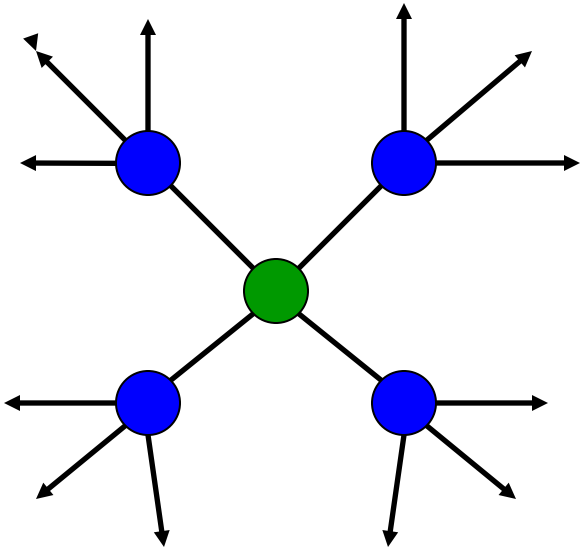
After k steps, reach d^k

When do we reach all n?

$d^k = n$ or $k = \log(n)/\log(d)$

Ideas:





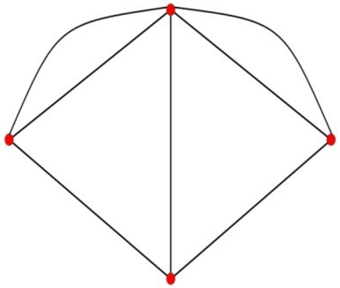
After k steps, reach d^k

When do we reach all n?

$d^k = n$ or $\mathbf{k = \log(n)/\log(d)}$

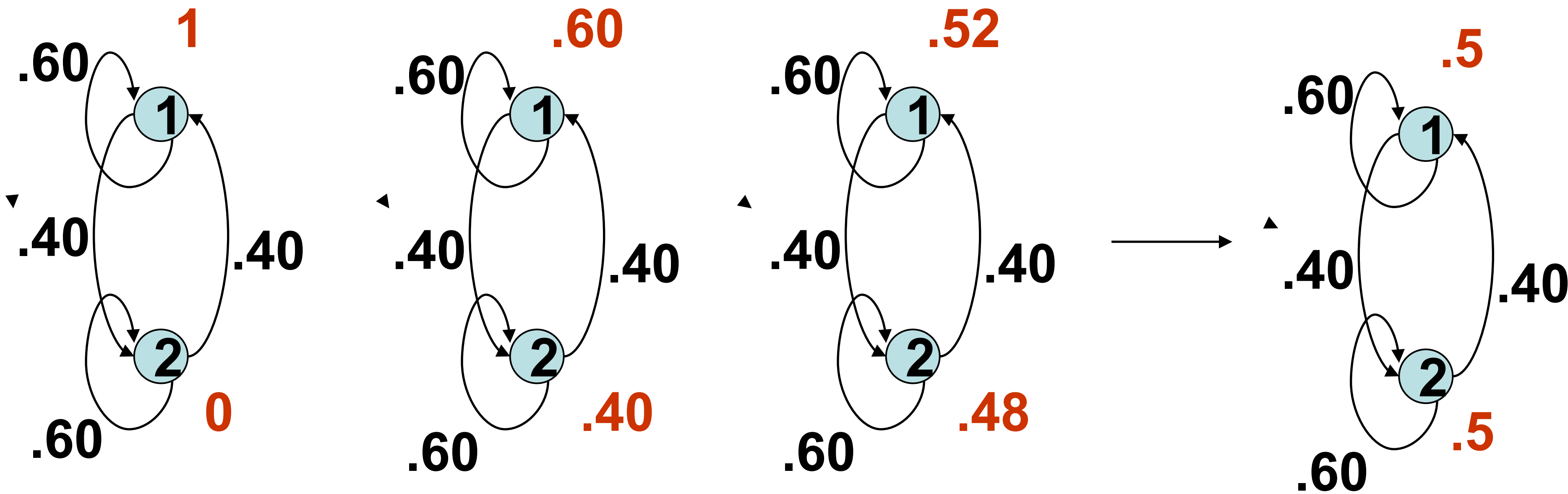
***Most* at maximum distance**
(10, 100, 1000, 10000...)

Small Worlds/Six Degrees of Separation

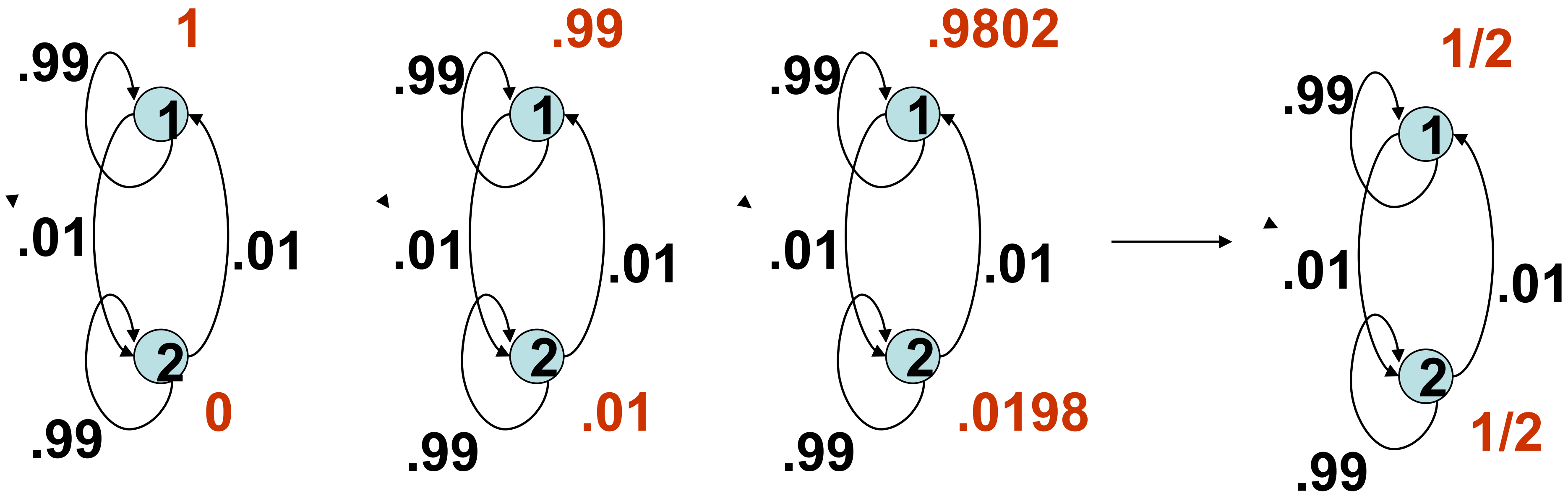


- n = 6.7 billion (world population)
- d = 50 (friends, relatives...)
- log(n)/log(d) is about 6 !!

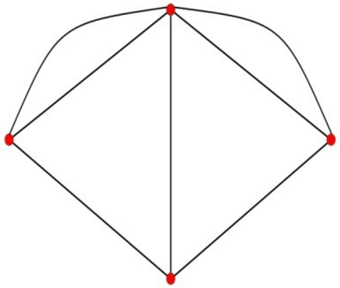
Learning Speed?



Learning Speed?



Measuring Homophily

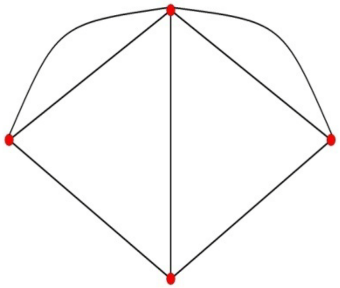


$$WH = (p_s - p_d) / (mp)$$

extra frequency of linking to own type relative to overall link frequency

[between -1 and 1]

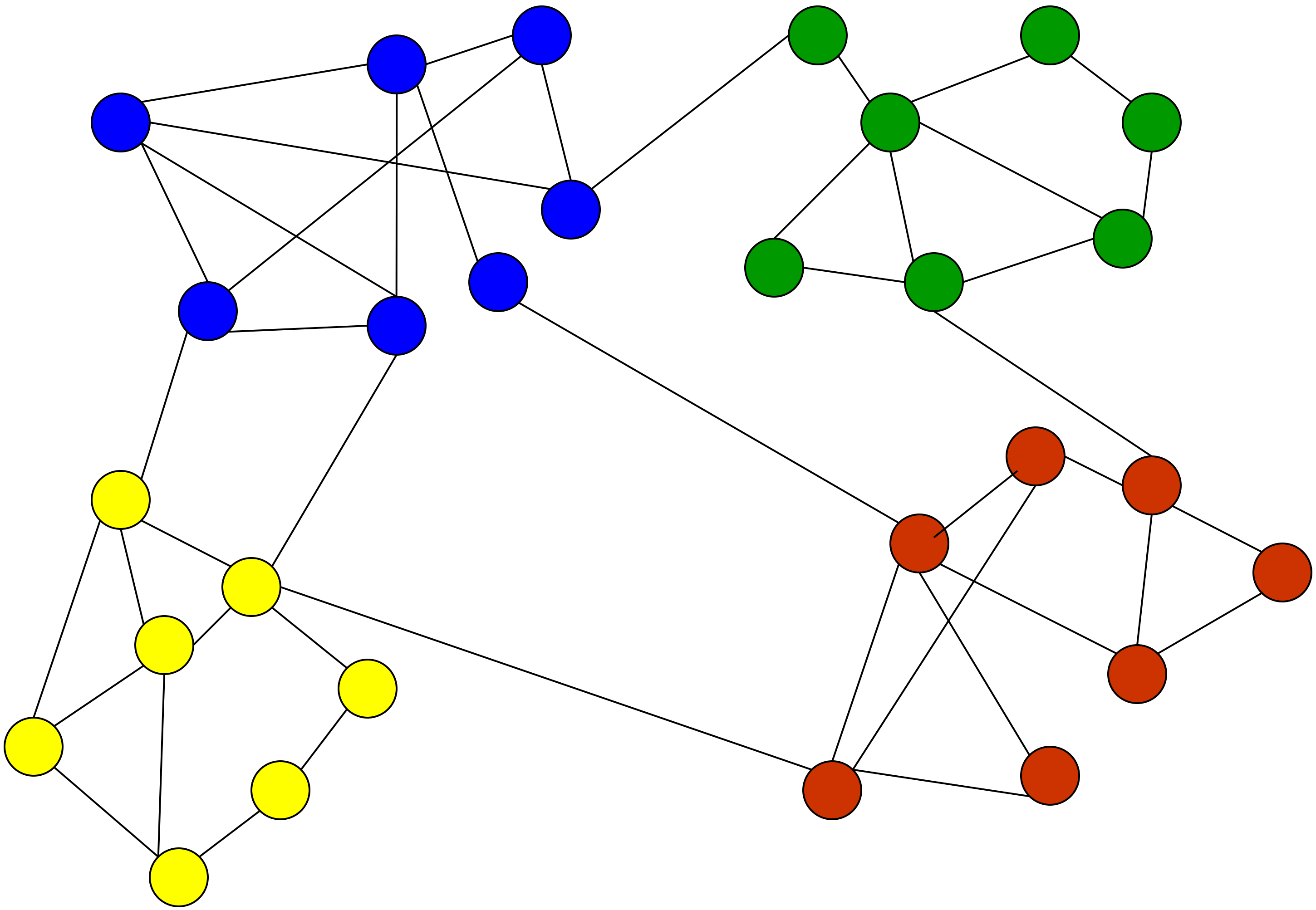
Theorem (Golub and Jackson 08)



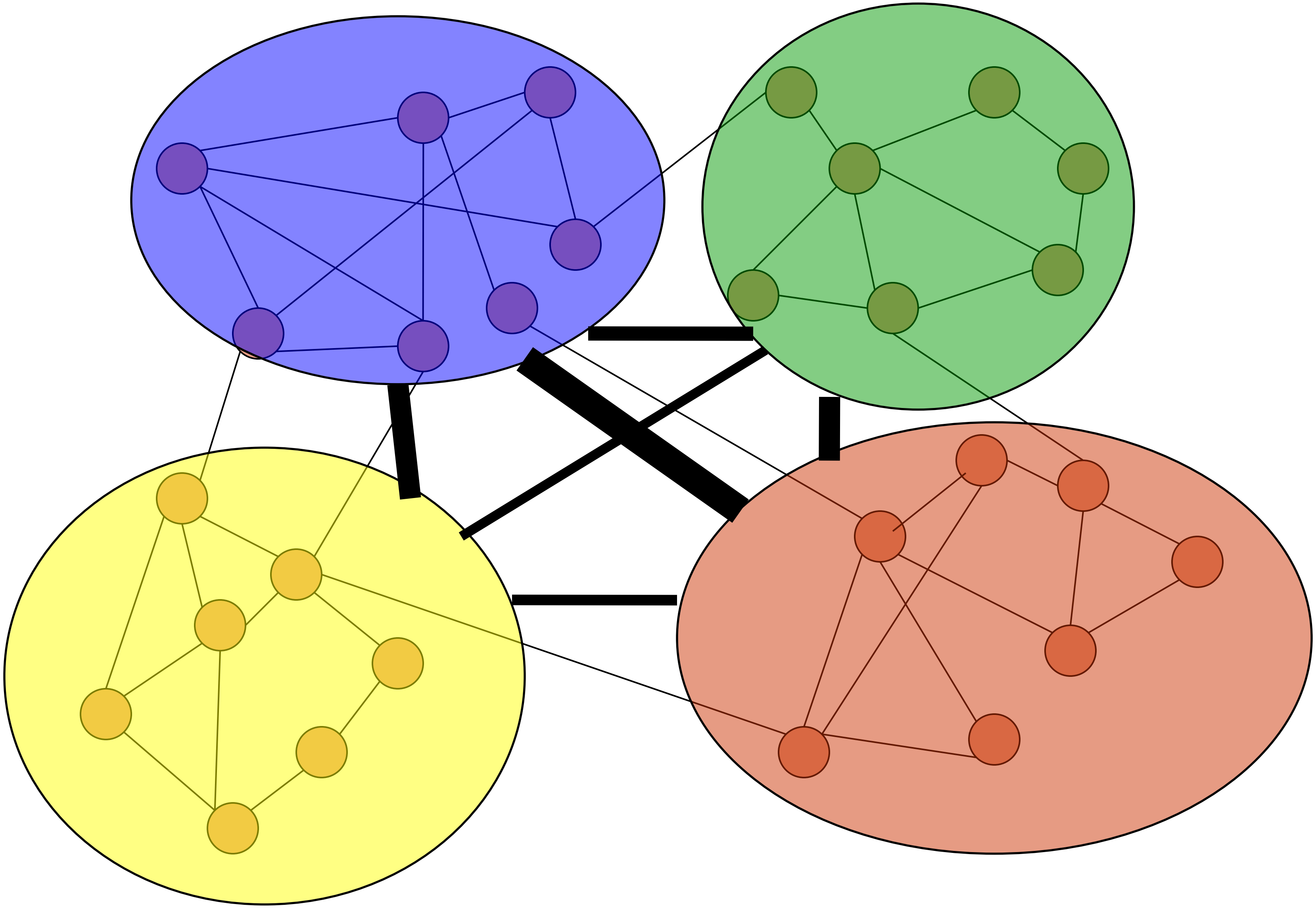
$$CT(T(n),1/n) \approx \log(n) / \log(1/|WH|)$$

[$\approx$ means Prob(within factor of 2) $\to_n 1$]

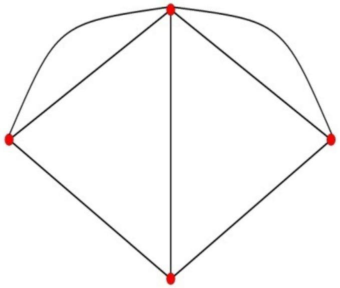
Representative Agents:

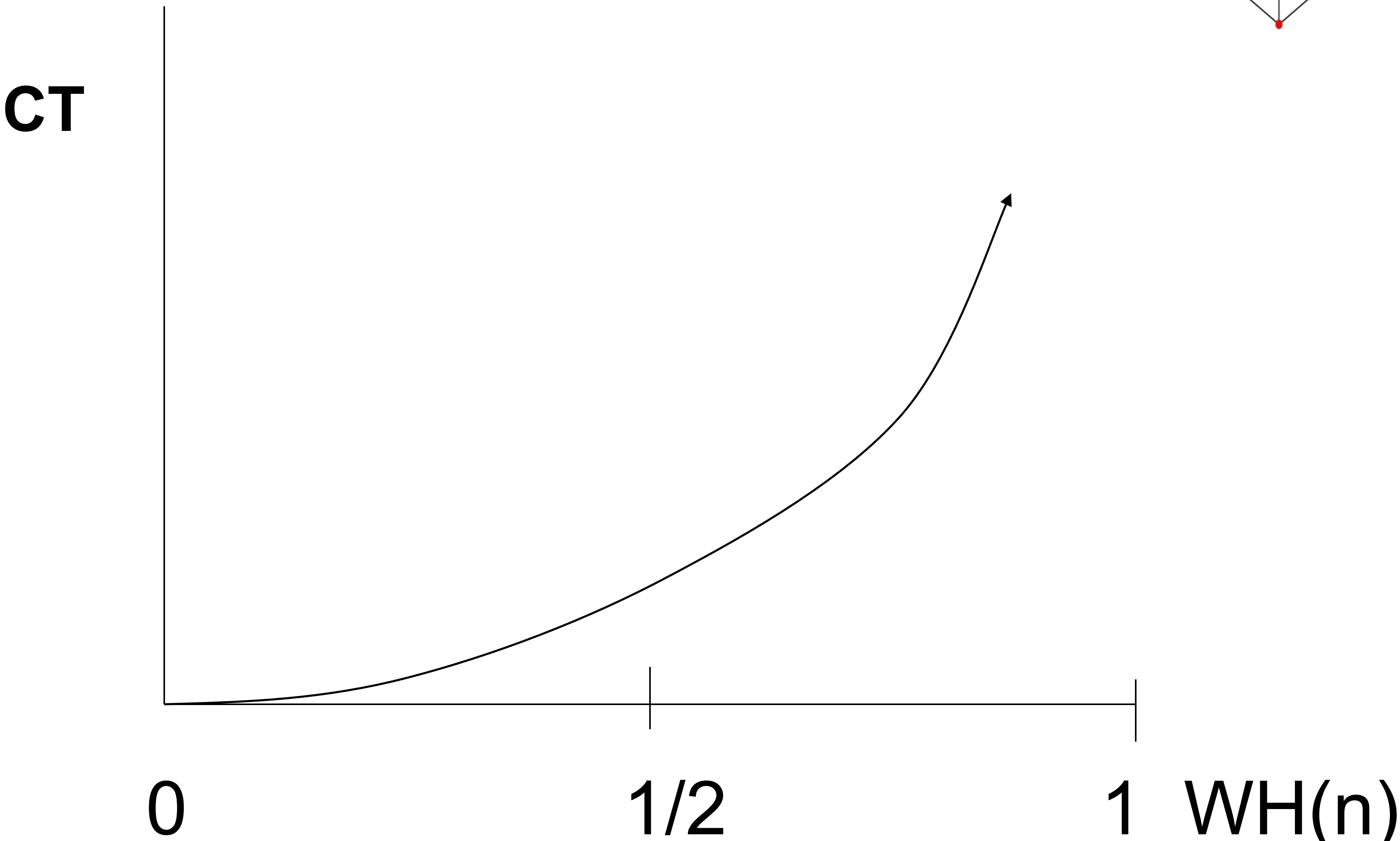


Representative Agents:

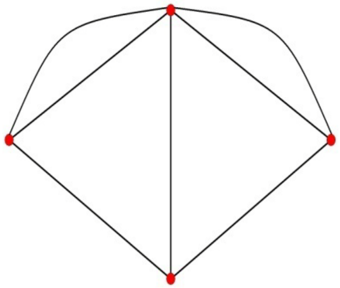


Consensus Time
Large n:





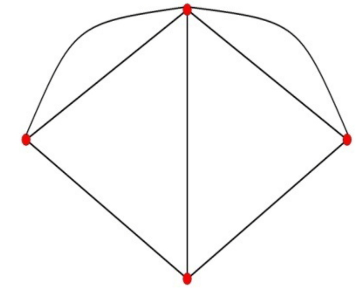
So



$$\text{AvgDist} \approx \log(n) / \log(d)$$

$$\text{CT} \approx \log(n) / \log(1/|WH|)$$

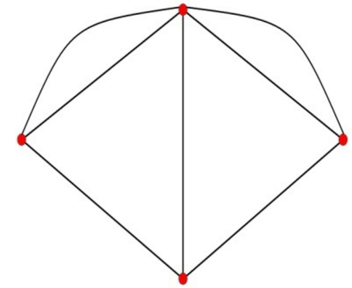
Speed:



Independent Variable

Process		Density	Homophily
	Shortest Path	↑	0
	Linear Updating	0	↓

Summary



- Pure Diffusion/Distance/Diameter is unaffected by segregation/homophily but is decreased by density
  - If the process on one link is unaffected by neighboring links, then network density matters but not the pattern

- Random Walks/Weighted Averaging/Updating is slowed by segregation/homophily but unaffected by density
  - If the process on one link is relative to the number of neighboring links then network pattern matters but not density