

Cascades in Networks and Agent Based Modeling

I. Makarov & L.E. Zhukov

BigData Academy MADE from Mail.ru Group

Network Science



Lecture outline

- 1 Diffusion of innovation
- 2 Influence propagation models
 - Independent cascade model
 - Linear threshold model
- 3 Influence maximization problem
 - Submodular function optimization
- 4 Spatial Segregation

Diffusion process

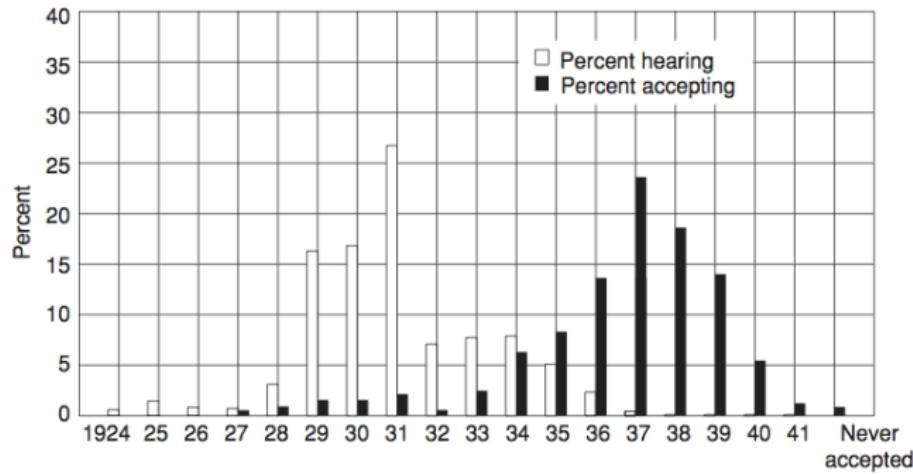
Propagation process:

- Viral propagation:
 - virus and infection
 - rumors, news
 - information
- Threshold (agent decision) models:
 - adoption of innovation
 - joining political protest
 - purchase decision
 - cascading failures

Local individual decision rules will lead to very different global results.
"microscopic" changes → "macroscopic" results

Diffusion of innovation

Ryan-Gross study of hybrid seed corn delayed adoption - diffusion of innovation

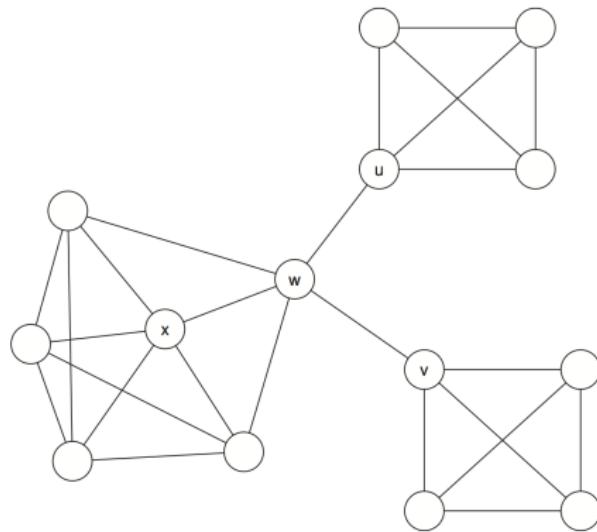


Information effect vs adopting of innovation

Ryan and Gross, 1943

Diffusion of innovation

Information (awareness) vs adoption (decision) spreading



Network coordination game

Local interaction game: Let u and v are players, and A and B are possible strategies

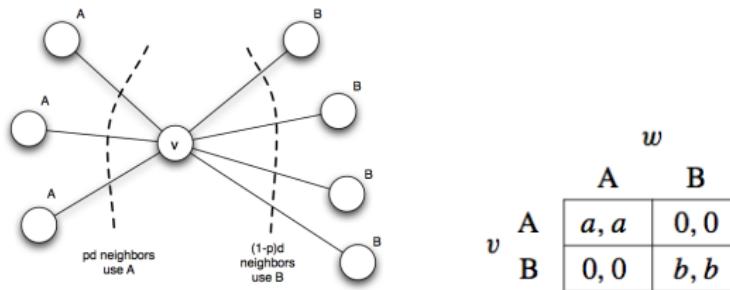
Payoffs

- if u and v both adopt behavior A, each get payoff $a > 0$
- if u and v both adopt behavior B, each get payoff $b > 0$
- if u and v adopt opposite behavior, each get payoff 0

		w
	A B	
v	A	a, a $0, 0$
	B	$0, 0$ b, b

Threshold model

Network coordination game, direct-benefit effect



Node v to make decision A or B , p - portion of type A neighbors to accept A :

$$a \cdot p \cdot d > b \cdot (1 - p) \cdot d$$

$$p \geq b/(a + b)$$

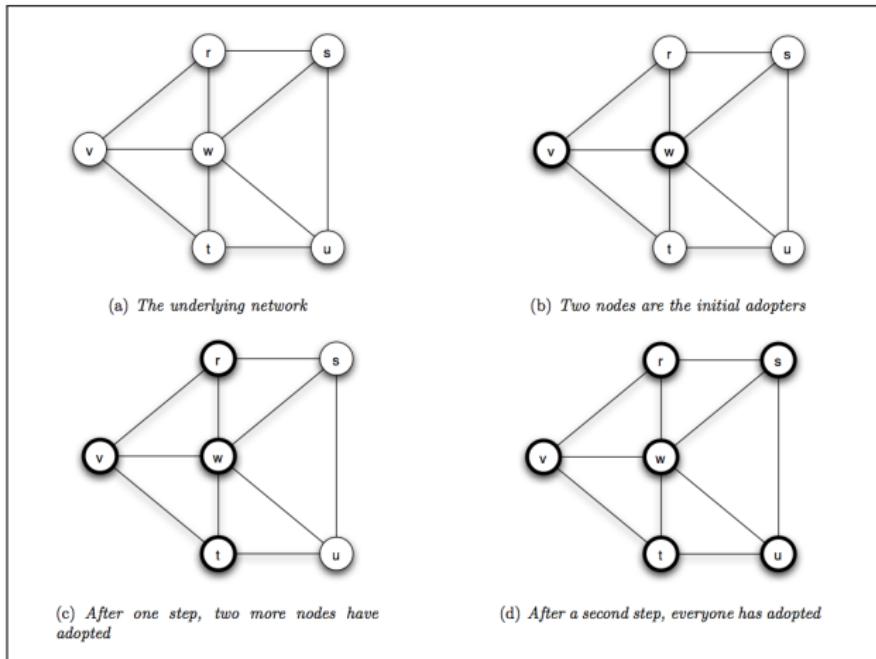
Threshold:

$$q = \frac{b}{a + b}$$

Accept new behavior A when $p \geq q$

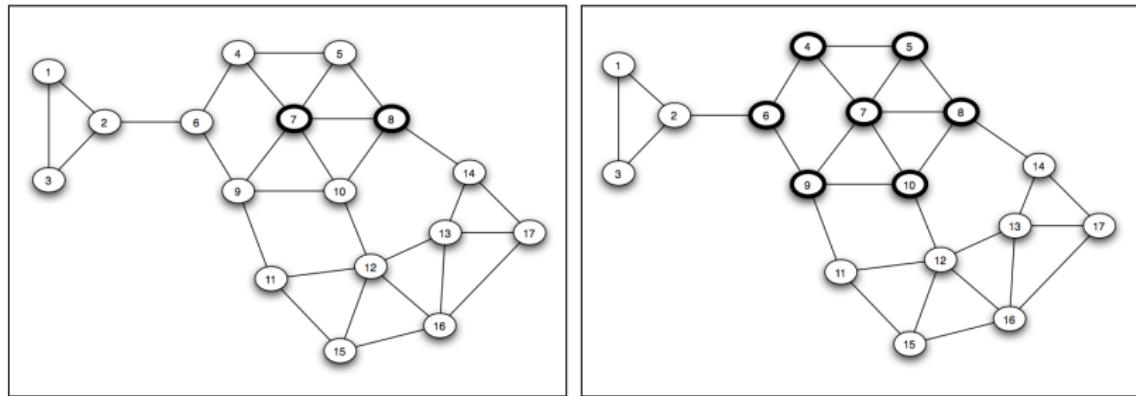
Cascades

Cascade - sequence of changes of behavior, "chain reaction"



Let $a = 3, b = 2$, threshold $q = 2/(2+3) = 2/5$

Cascade propagation

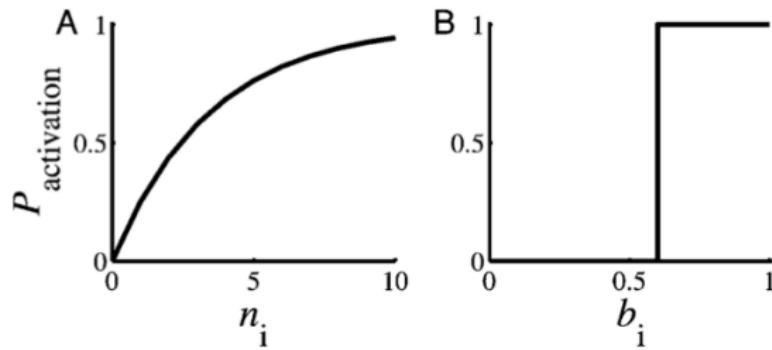


- Let $a = 3$, $b = 2$, threshold $q = 2/(2+3) = 2/5$
- Start from nodes 7,8: $1/3 < 2/5 < 1/2 < 2/3$
- Cascade size - number of nodes that changed the behavior
- Complete cascade when every node changes the behavior

Influence response

Two models:

- Independent Cascade Model (diminishing returns)
- Linear Threshold Model (critical mass)



$$P(n) = 1 - (1 - p)^n \quad P(b) = \delta(b > b_0)$$

Influence response: diminishing returns and threshold

D. Kempe, J. Kleinberg, E. Tardos, 2003

Independent cascade model

- Initial set of active nodes S_0
- Discrete time steps
- On every step an active node v can activate connected neighbor w with a probability $p_{v,w}$ (single chance)
- If v succeeds, w becomes active on the next time step
- Process runs until no more activations possible

If $p_{v,w} = p$ it is a particular type of SIR model, a node stays infected for only one step

D. Kempe, J. Kleinberg, E. Tardos, 2003

Linear threshold model

- Influence comes only from NN $N(i)$ nodes, w_{ij} influence $i \rightarrow j$
- Require $\sum_{j \in N(i)} w_{ji} \leq 1$
- Each node has a random acceptance threshold from $\theta_i \in [0, 1]$
- Activation: fraction of active nodes exceeds threshold

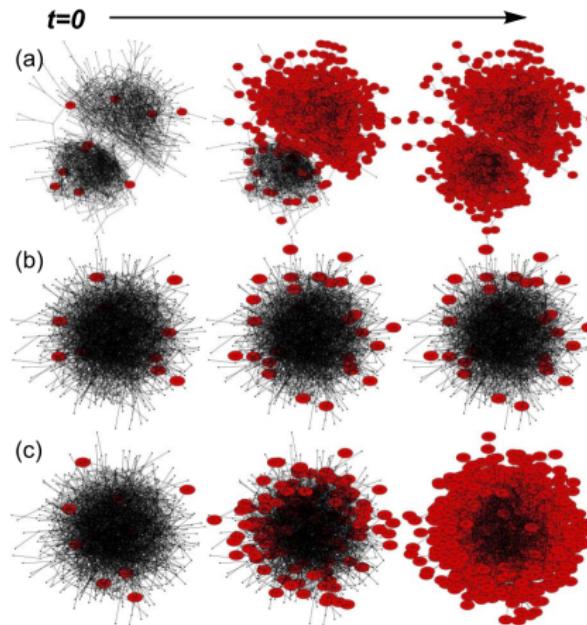
$$\sum_{\substack{\text{active } \\ j \in N(i)}} w_{ji} > \theta_i$$

- Initial set of active nodes A_o , iterative process with discrete time steps
- Progressive process, only nonactive \rightarrow active

D. Kempe, J. Kleinberg, E. Tardos, 2003

Cascades in random networks

multiple seed nodes



(a) Empirical network; (b), (c) - randomized network

P. Singh, 2013

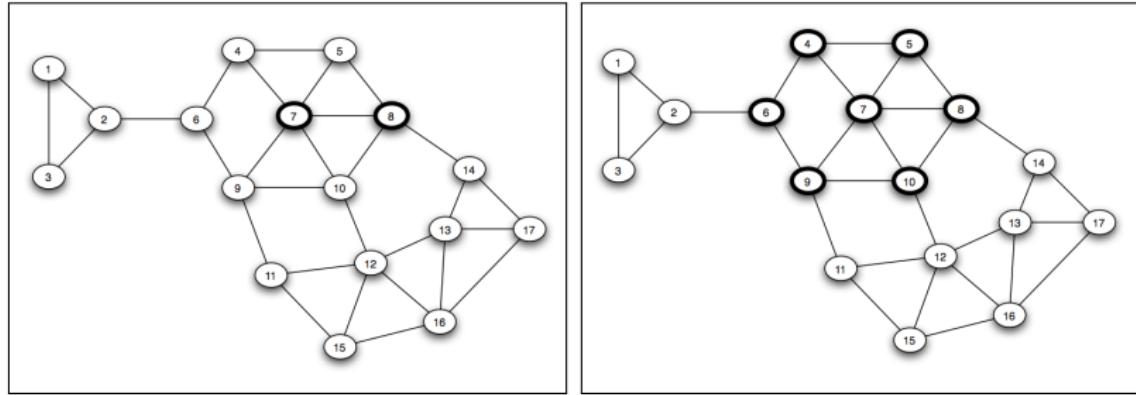
I. Makarov & L.E. Zhukov (MADE)

Lecture 7

05.04.2021

13 / 42

Influence maximization problem



- Initial set of active nodes A_o
- Cascade size $\sigma(A_o)$ - expected number of active nodes when propagation stops
- Find k -set of nodes A_o that produces maximal cascade $\sigma(A_o)$
- k -set of "maximum influence" nodes
- NP-hard

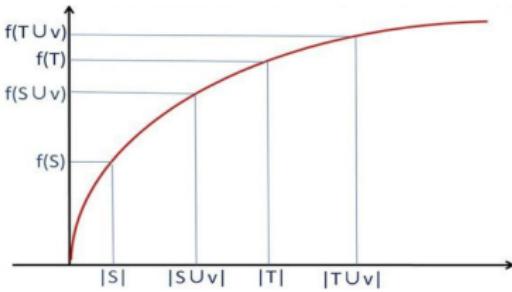
D. Kempe, J. Kleinberg, E. Tardos, 2003, 2005

Submodular functions

- Set function f is submodular, if for sets S, T and $S \subseteq T, \forall v \notin T$

$$f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$$

- Function of diminishing returns ("concave property")
- Function f is monotone, $f(S \cup \{v\}) \geq f(S)$



Submodular functions

Theorem

Let F be a monotone submodular function and let S^* be the k -element set achieving maximal f .

Let S be a k -element set obtained by repeatedly, for k -iterations, including an element producing the largest marginal increase in f .

$$f(S) \geq \left(1 - \frac{1}{e}\right)f(S^*)$$

Nemhauser, Wolsey, and Fisher, 1978

Influence maximization

- Cascade size $\sigma(S)$ is submodular function (D. Kempe, J. Kleinberg, E. Tardos, 1993)

$$\sigma(S) \geq \left(1 - \frac{1}{e}\right)\sigma(S^*)$$

- Greedy algorithm for maximum influence set finds a set S such that its influence set $\sigma(S)$ is within $1/e = 0.367$ from the optimal (maximal) set $\sigma(S^*)$, $\sigma(S) \geq 0.629\sigma(S^*)$

Influence maximization

Approximation algorithm

Algorithm: Greedy optimization

Input: Graph $G(V, E)$, k

Output: Maximum influence set S

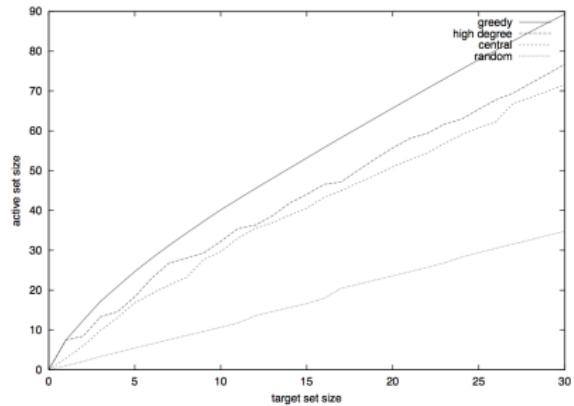
Set $S \leftarrow \emptyset$

for $i = 1 : k$ **do**

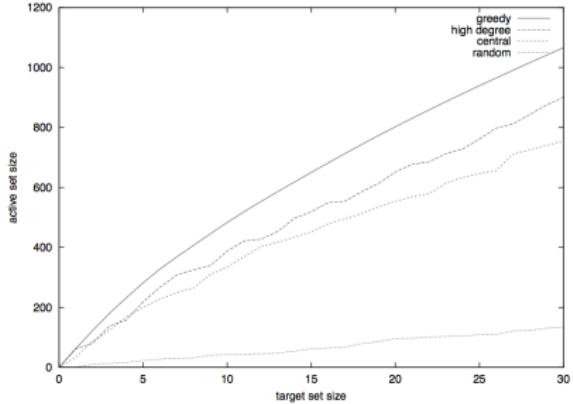
select $v = \arg \max_{u \in V \setminus S} (\sigma(S \cup \{u\}) - \sigma(S))$

$S \leftarrow S \cup \{v\}$

Experimental results



Independent cascade model

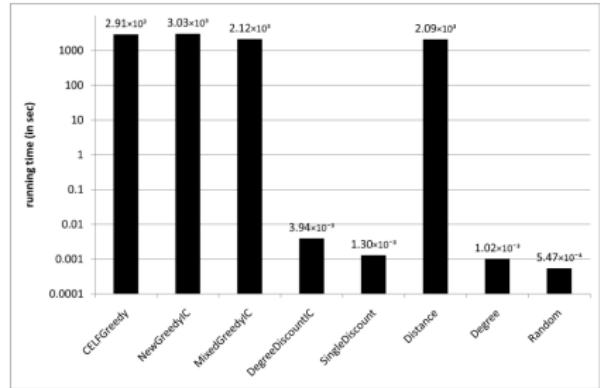
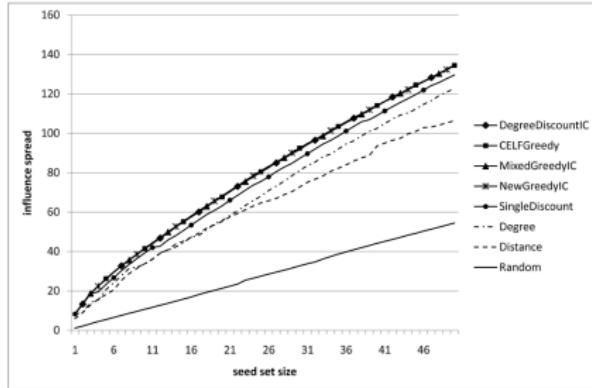


Linear threshold model

network: collaboration graph 10,000 nodes, 53,000 edges

D. Kempe, J. Kleinberg, E. Tardos, 2003

Computational considerations



Independent cascade model: influence spread and running time

W. Chen et.al, 2009

Lecture outline

- 1 Diffusion of innovation
- 2 Influence propagation models
 - Independent cascade model
 - Linear threshold model
- 3 Influence maximization problem
 - Submodular function optimization
- 4 Spatial Segregation

Agent Based Modeling (ABM)

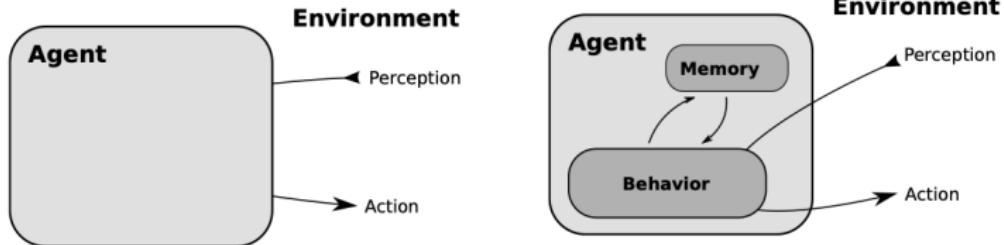
An agent-based model (ABM) is a class of computational models for simulating the actions and interactions of autonomous agents

- Agent-based models consist of dynamically interacting rule-based agents.
- Simple agent behaviors (rules) generate complex system behaviors
- Real world system becoming very complex and interdependent
- Decentralization of decision making, deregulations.
- Available data and computational power for micro simulations

What is an agent?

Agents are autonomous decision-making units with diverse characteristics

- A discrete entity with its own goals and behavior
- Autonomous, capable to adapt and modify behavior
- Decisions made independently by each engine
- Agents can be homogeneous or diverse and heterogeneous
- Can have memory and internal models
- Examples: people, groups, organizations, systems of robots etc



Agent based simulations

- Agent based model consists of a set of user defined agents, a set of agent relationships and environment
- No central controller or authority exists for the system
- Independent move and interaction by any agent
- Local interaction among agents
- Various topologies connect agents with their neighbors (fee space, grid, network, GIS)
- Optimization can be done for the system globally

When use agent based modeling?

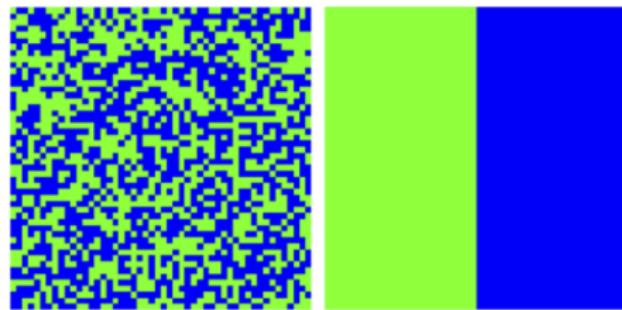
- When there is a natural representation as agents
 - When there are decision and behaviors that can be defined discretely (with boundaries)
 - When it is important that agents adapt and change their behavior
 - When it is important that agents learn and engage in dynamic strategic behavior
 - When it is important that agents have a dynamic relationships with other agents, and agent relationships form and dissolve
 - When it is important that agents have a spatial component to their behaviors and interactions
- When the past is no predictor of the future
- When scale-up to arbitrary levels is important
- When process structural change needs to be a result of the model, rather than an input to the model

Spatial model of segregation

"Dynamic Models of Segregation", Thomas Schelling, 1971

- Micro-motives and macro-behavior
- Personal preferences lead to collective actions
- Global patterns of spatial segregation from homophily at a local level
- Segregated race, ethnicity, native language, income
- Cities are strongly racially segregated. Are people that racists?
- Agent based modeling: agents, rules (dynamics), aggregation

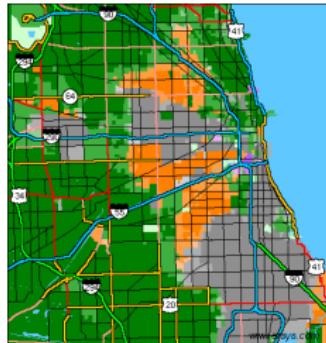
Segregation



Integrated pattern

Segregated pattern

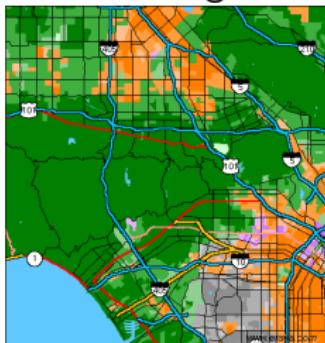
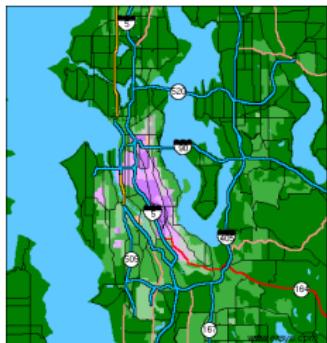
Racial segregation



New York

Washington

Chicago

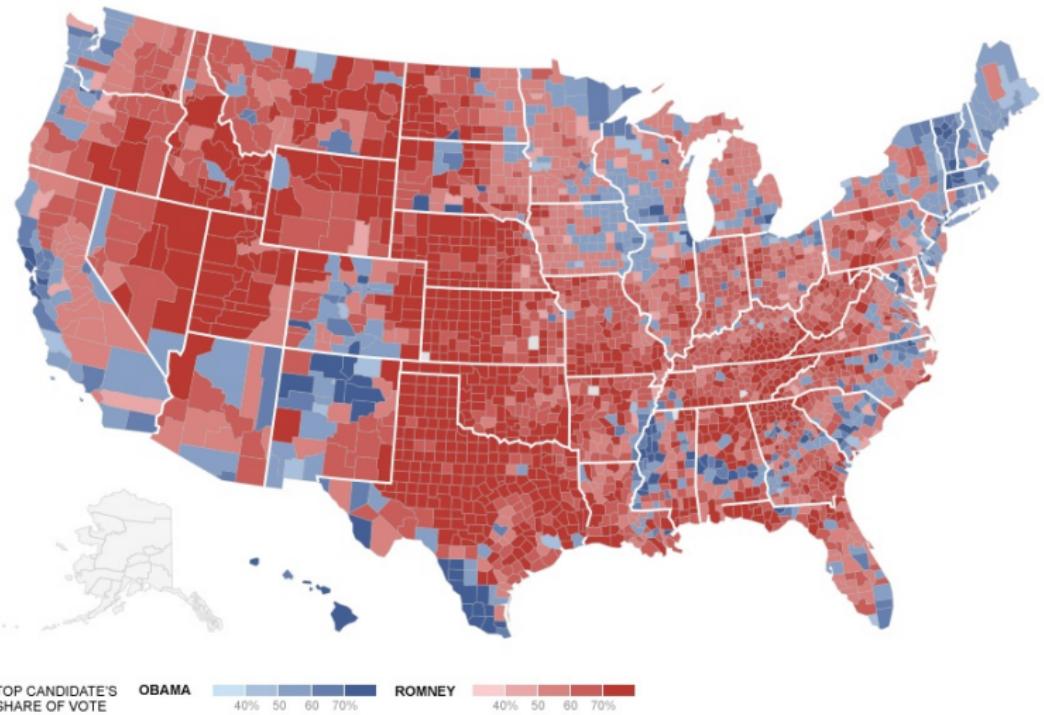


Seattle

Los Angeles

Miami

2012 US Presidential Elections Map



Schelling's model of segregation

- Population consists of 2 types of agents
- Agents reside in the cells of the grid (2-dimensional geography of a city), 8 neighbors
- Some cells contain agents, some unpopulated
- Every agent wants to have at least some fraction of agents (threshold) of his type as neighbor (satisfied agent)
- On every round every unsatisfied agent moves to a satisfactory empty cell.
- Continues until everyone is satisfied or can't move

Spatial segregation

1	2	3
4	X	5
6	7	8

satisfied agent

1	2	3
4	X	5
6	7	8

unsatisfied agent

- preference threshold $\lambda = 3/7$

Model

- N - nodes, θ - fraction of occupied by A and B

$$n_A + n_B = \theta \cdot N$$

- Proportion of "unlike" nearest neighbors, $k_i = \#\text{NN}$

$$P_i = \begin{cases} \#n_B/k_i, & \text{if } i \in A \\ \#n_A/k_i, & \text{if } i \in B \end{cases}$$

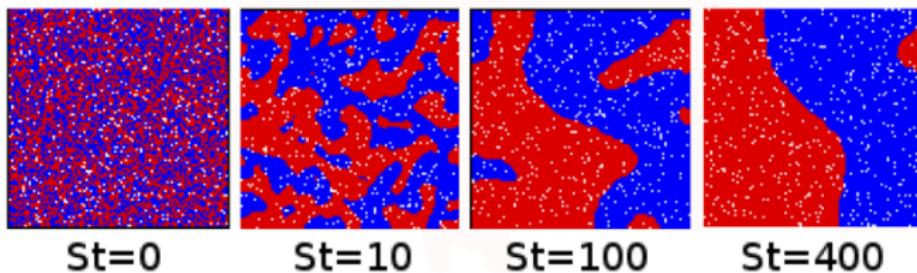
- Utility function, λ - sensitivity (tolerance threshold) level

$$u_i = \begin{cases} 1, & \text{if } P_i \leq \lambda \\ 0, & \text{if } P_i > \lambda \end{cases}$$

- Every node moves to maximize its utility

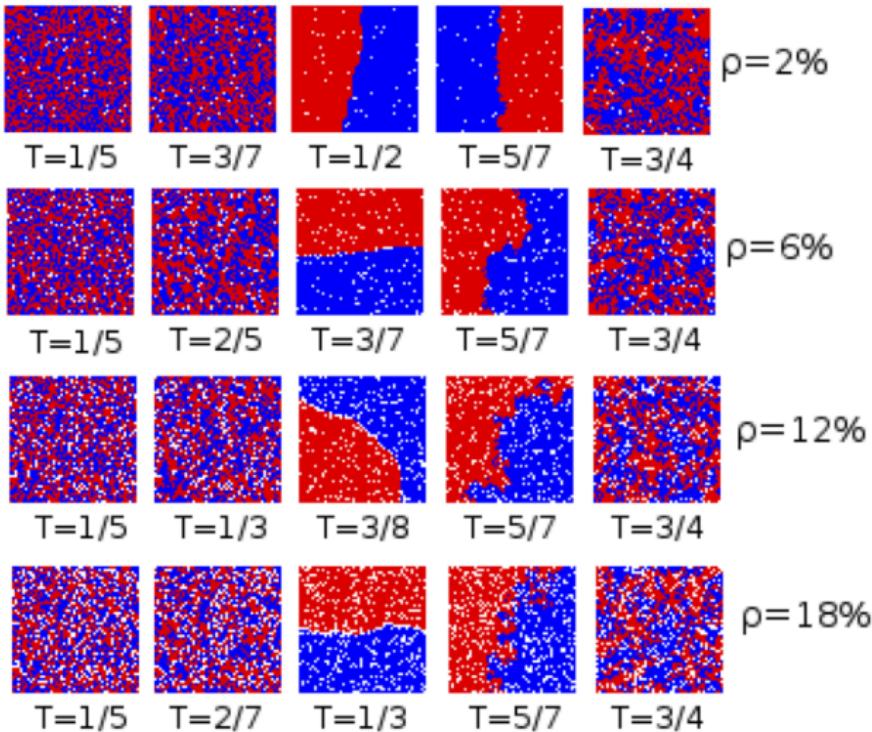
Spatial segregation

vacancy 5%, tolerance $\lambda = 0.5$



L. Gauvin et.al. 2009

Spatial segregation



Algorithm

- time steps $1..T$
- At every time step randomly select an agent, compute utility
- If utility is $u = 0$ move to an empty location to maximize utility
- Movements: 1) random location 2) nearest available location
- Repeat until either all utilities are maximized $\sum_i u_i = \theta N$ or reaches "frozen" state, no place to move, then $\sum_i u_i < \theta N$
- Total utility of society

$$U = \sum_i u_i$$

Measuring segregation

- Schilling's solid mixing index

$$M = \frac{1}{n_A + n_B} \sum_i P_i$$

- Freeman's segregation index

$$F = 1 - \frac{e^*}{E(e^*)}$$

$e^* = \frac{e_{AB}}{(e_{AB} + e_{AA} + e_{BB})}$ - observed proportion of between group ties,

$E(e^*) = \frac{2n_A n_B}{(n_A + n_B)(n_A + n_B - 1)}$ - expected proportion for random ties

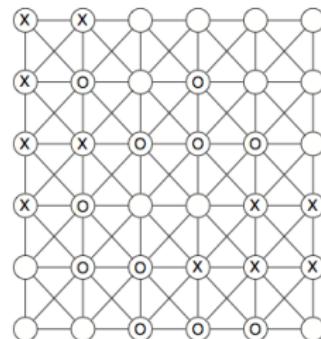
- Assortative mixing

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j)$$

Spatial segregation on networks

X	X					
X	O			O		
X	X	O	O	O		
X	O				X	X
	O	O	X	X	X	X
		O	O	O		

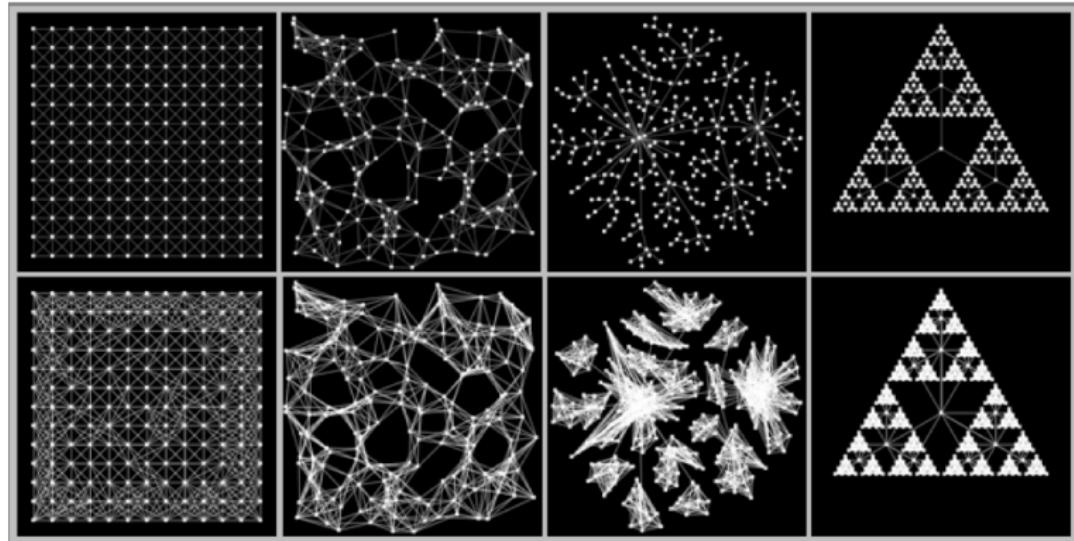
(a)



(b)

Spatial segregation on networks

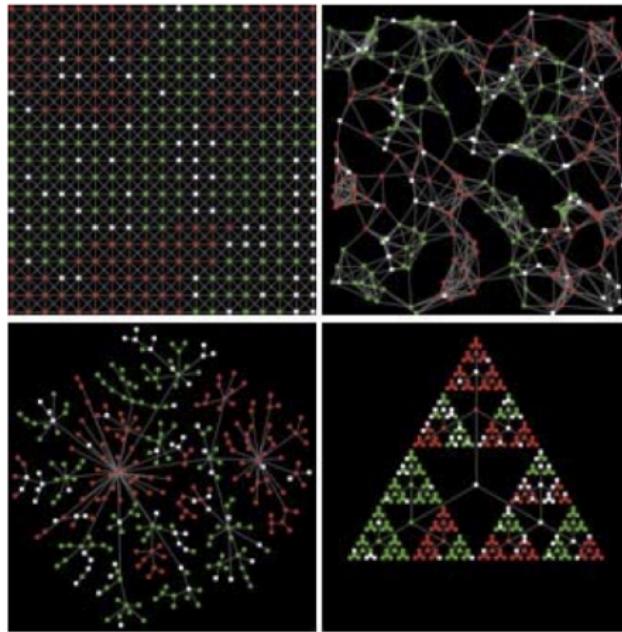
Fixed degree $k = 10$ neighboring graphs: regular, random, scale-free, fractal



Arnaud Banos, 2010

Spatial segregation on networks

$$\lambda = 0.5, \theta = 0.8$$



Banos, 2010

Summary

- Spatial segregation is taking place even though no individual agent is actively seeking it (minor preferences, high tolerance)
- Network structure does affect segregation
- Fixed characteristics (race) can become correlated with mutable (location)

References

- Contagion, S. Morris, Review of Economic Studies, 67, p 57-78, 2000
- Maximizing the Spread of Influence through a Social Network, D. Kempe, J. Kleinberg, E. Tardos, 2003
- Influential Nodes in a Diffusion Model for Social Networks, D. Kempe, J. Kleinberg, E. Tardos, 2005
- Efficient Influence Maximization in Social Networks, W. Chen, Y. Wang, S. Yang, KDD 2009.
- A Simple Model of Global Cascades on Random Networks. D. Watts, 2002.

References

- Dynamic Models of Segregation, Thomas C. Schelling, 1971
- Segregation in Social Networks, Linton Freeman, 1978
- Gauvin L, Vannimenus J, Nadal JP. Phase diagram of a Schelling segregation model. *The European Physical Journal B*, 70:293-304, 2009
- Arnaud Banos. Network effects in Schelling's model of segregation: new evidences from agent-based simulations. 2010