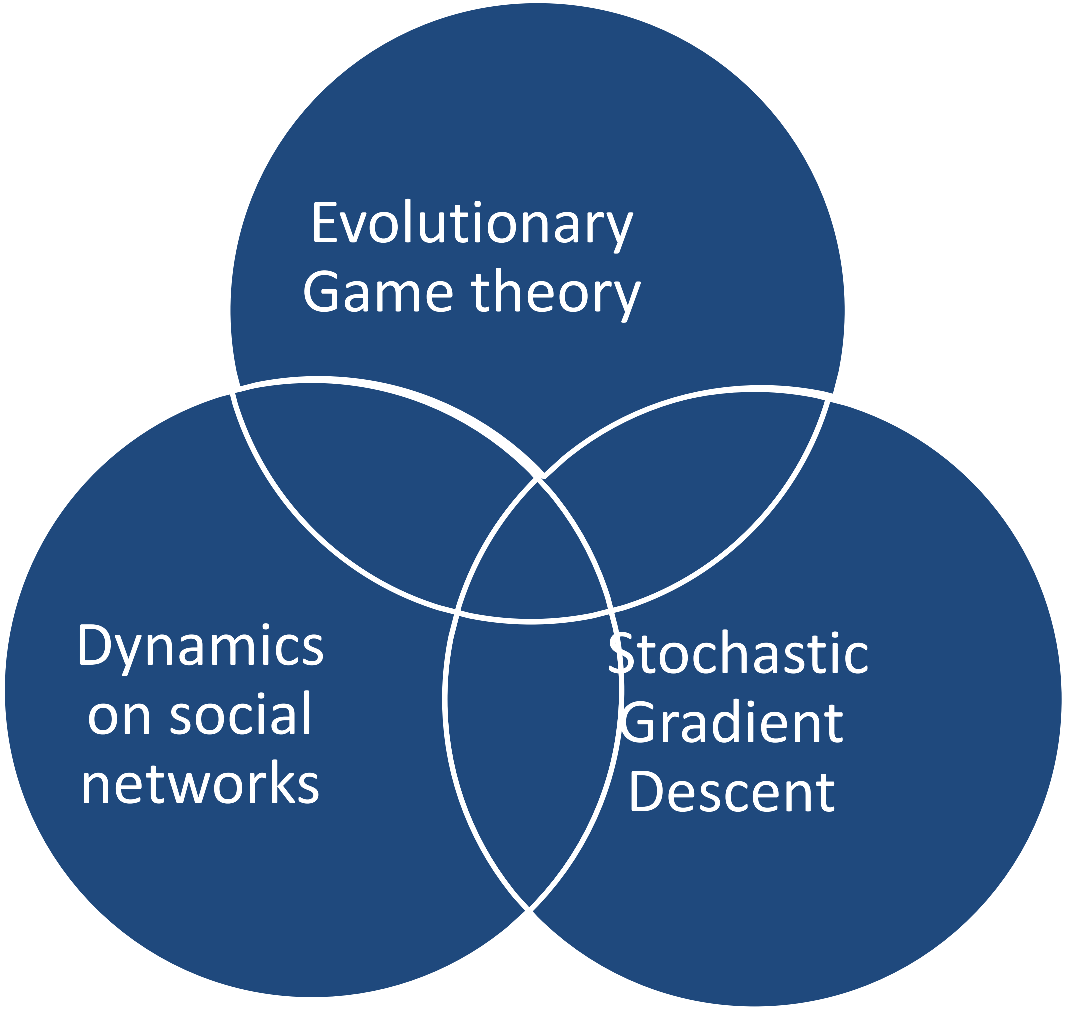
Escaping Saddle Points in Constant Dimensional Spaces: an Agent-based Modeling Perspective

Grant Schoenebeck, University of Michigan

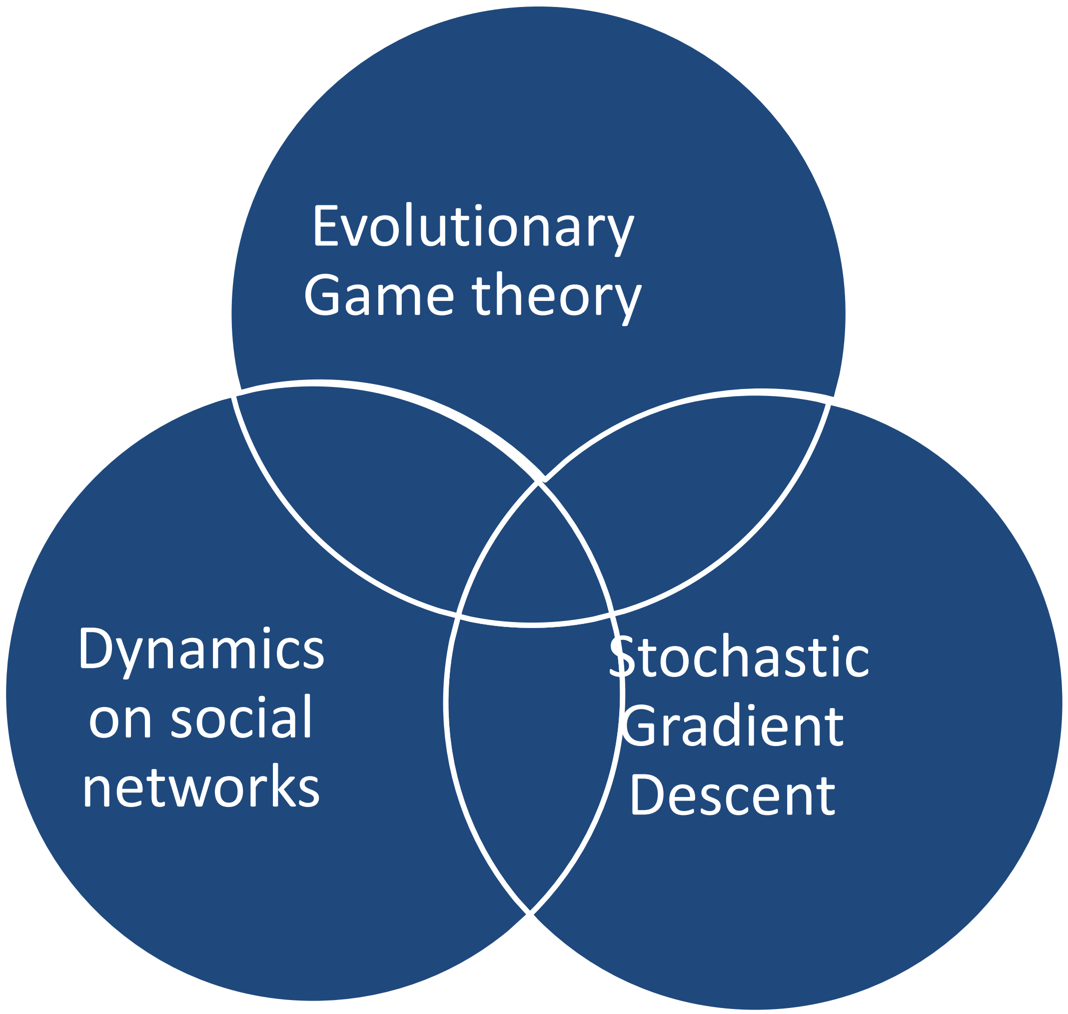
Fang-Yi Yu, Harvard University

Results

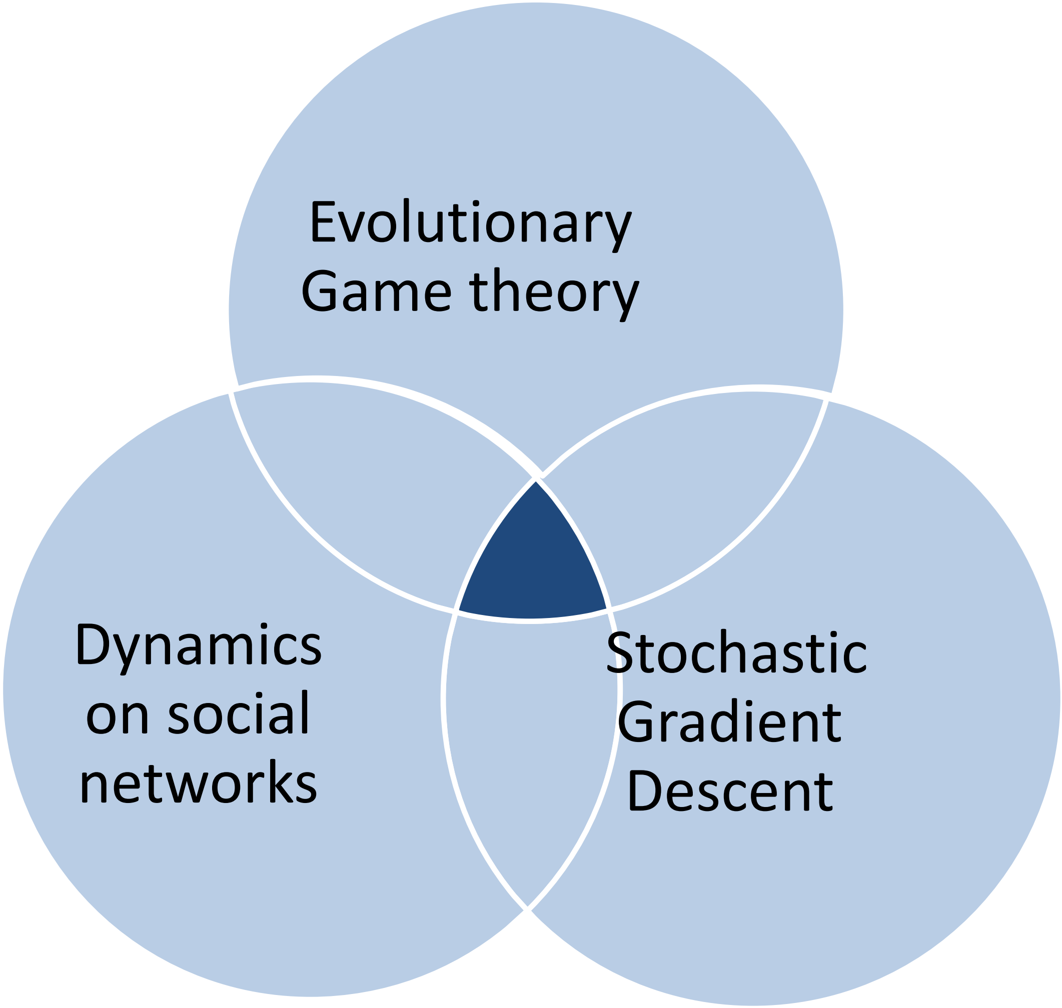
- Analyze the convergence rate of a family of stochastic processes
- Three related applications
  - Evolutionary game theory
  - Dynamics on social networks
  - Stochastic Gradient Descent



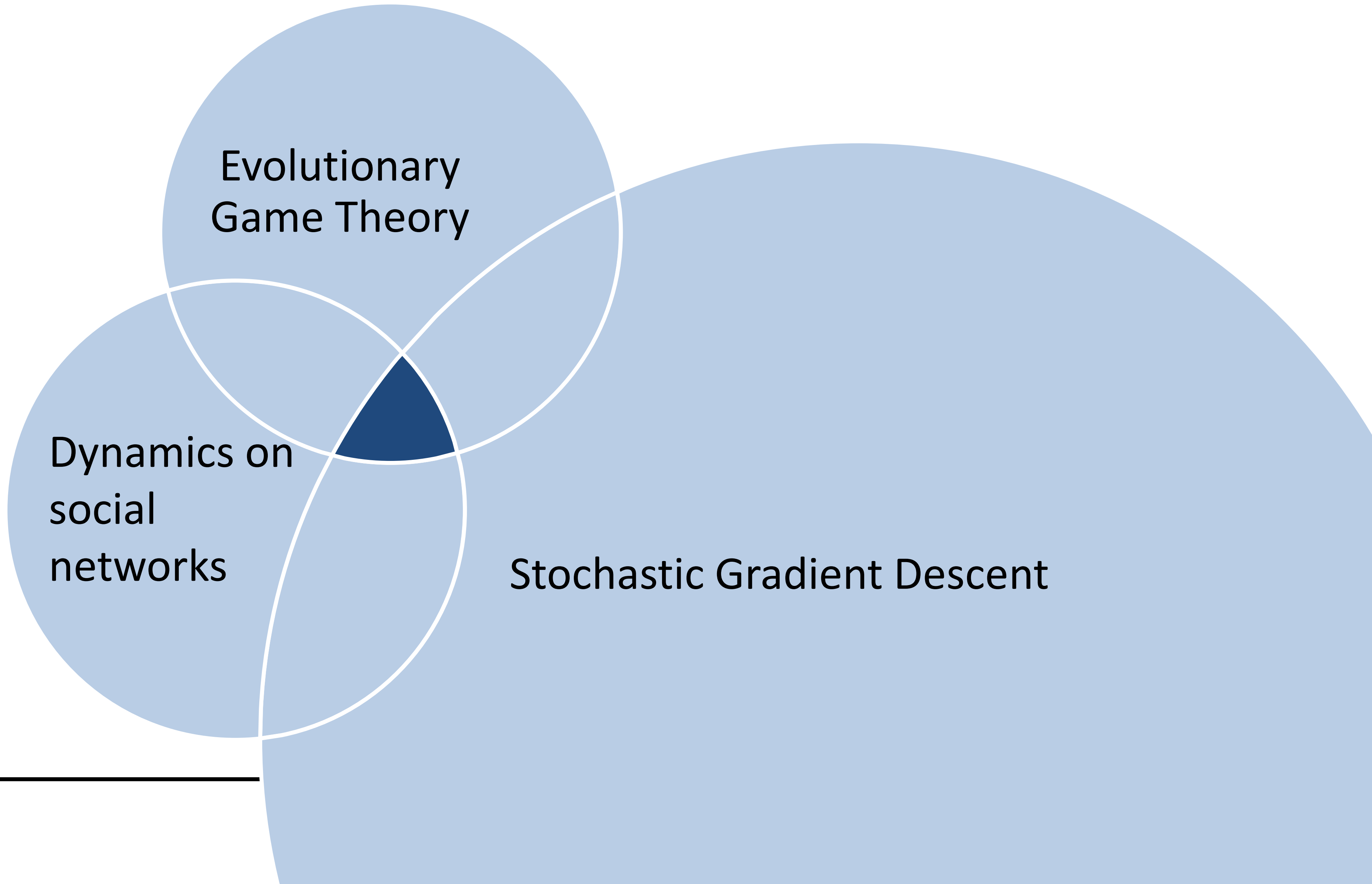
Target Audience



Target Audience

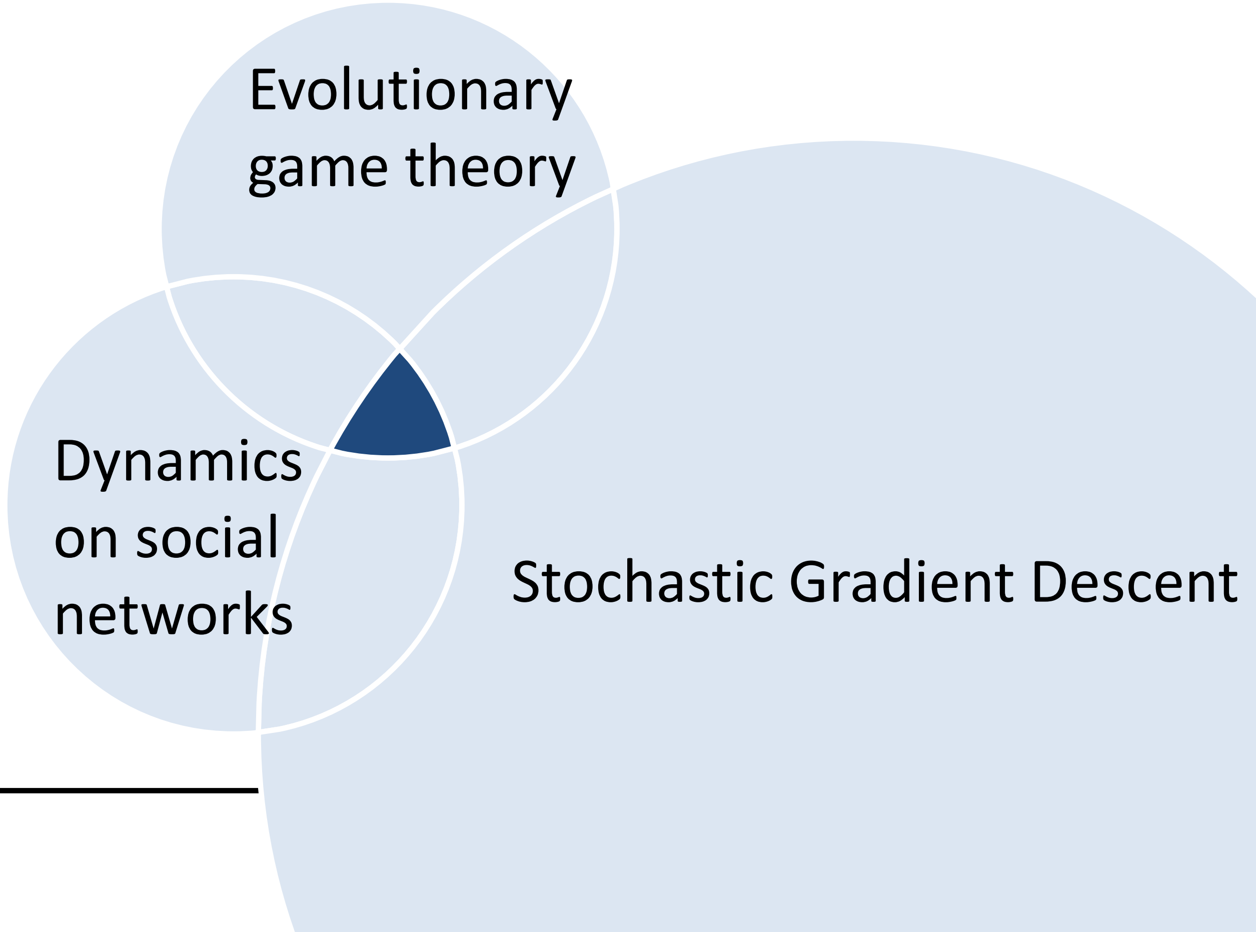


Target Audience (still not-to-scale)



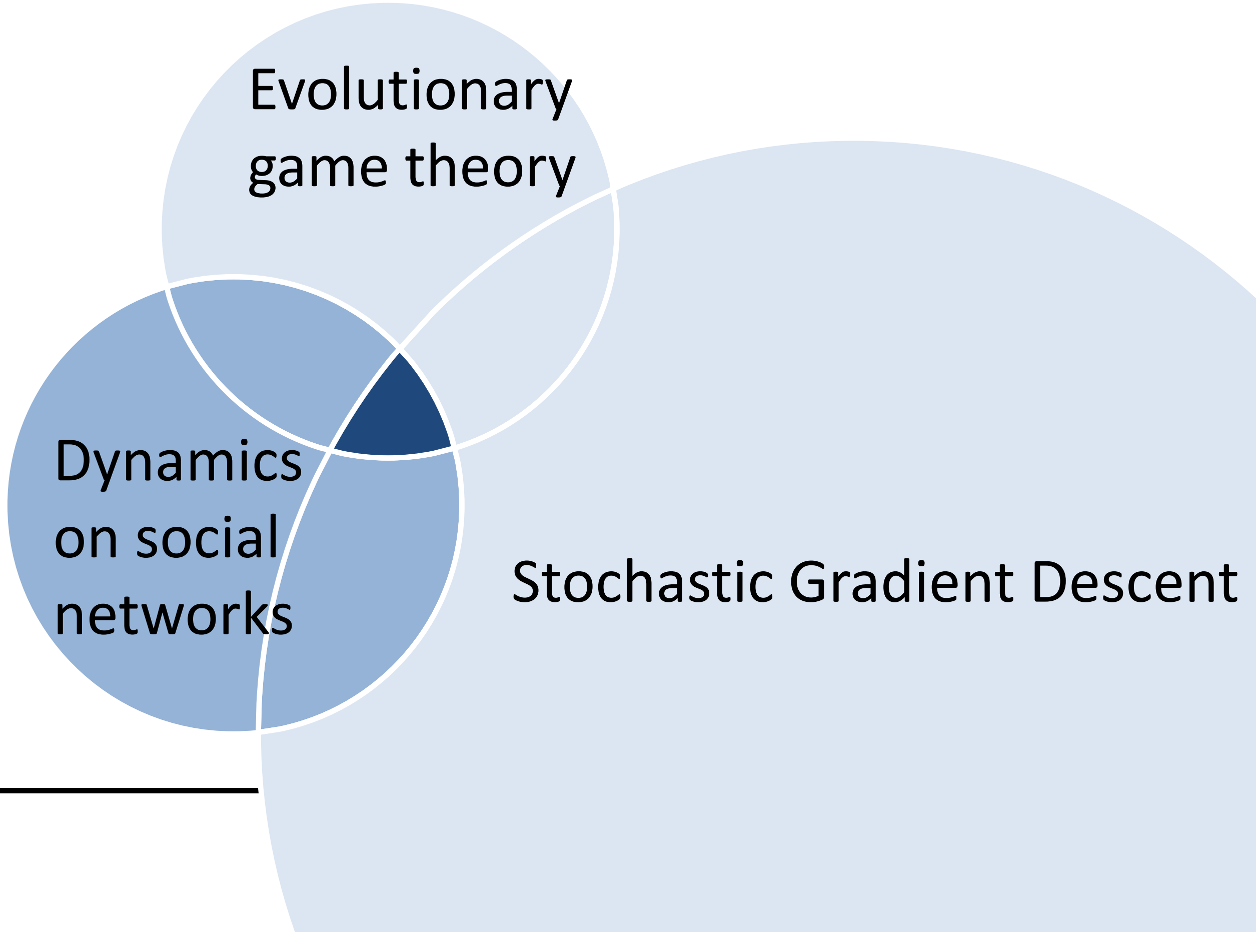
Outline

- Escaping saddle point



Outline

- Escaping saddle point
- Case study: dynamics on social networks



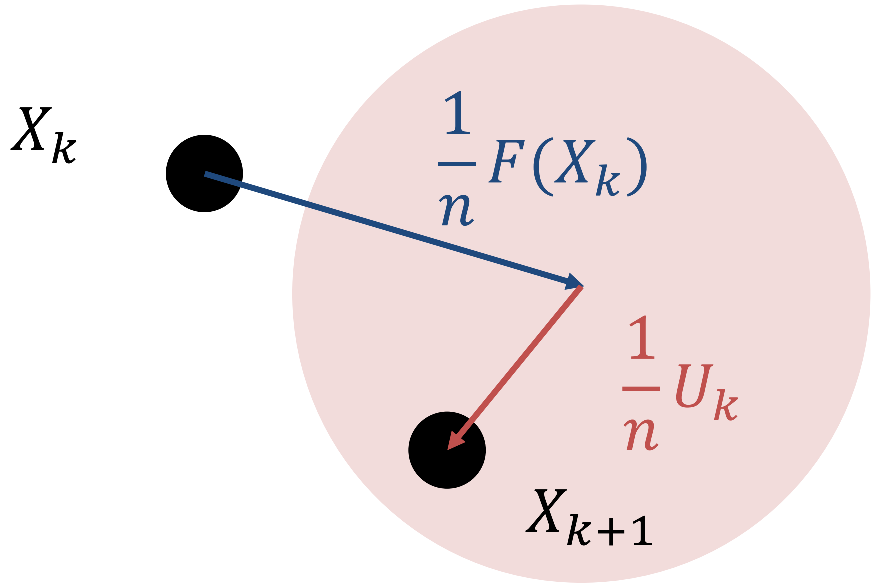
Upper bounds and lower bounds

ESCAPING SADDLE POINTS

Reinforced random walk with F

A discrete time stochastic process $\{X_k : k = 0, 1, \ldots\}$ in $\mathbb{R}^d$ that admits the following representation,

$$X_{k+1} - X_k = \frac{1}{n}(F(X_k) + U_k)$$

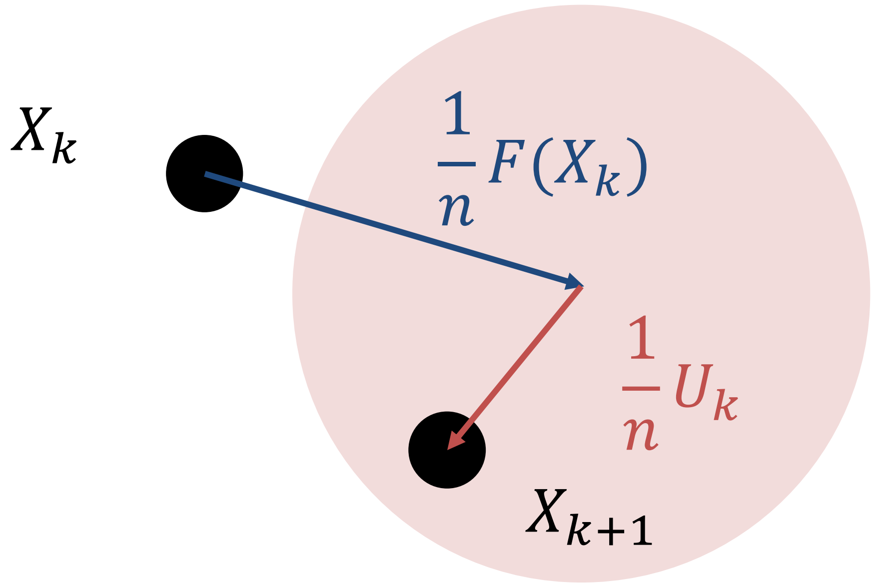


Reinforced random walk with F

A discrete time stochastic process $\{X_k : k = 0, 1, \dots\}$ in $\mathbb{R}^d$ that admits the following representation,

$$X_{k+1} - X_k = \frac{1}{n}\left(F(X_k) + U_k\right)$$

- Expected difference (drift), $F(X)$

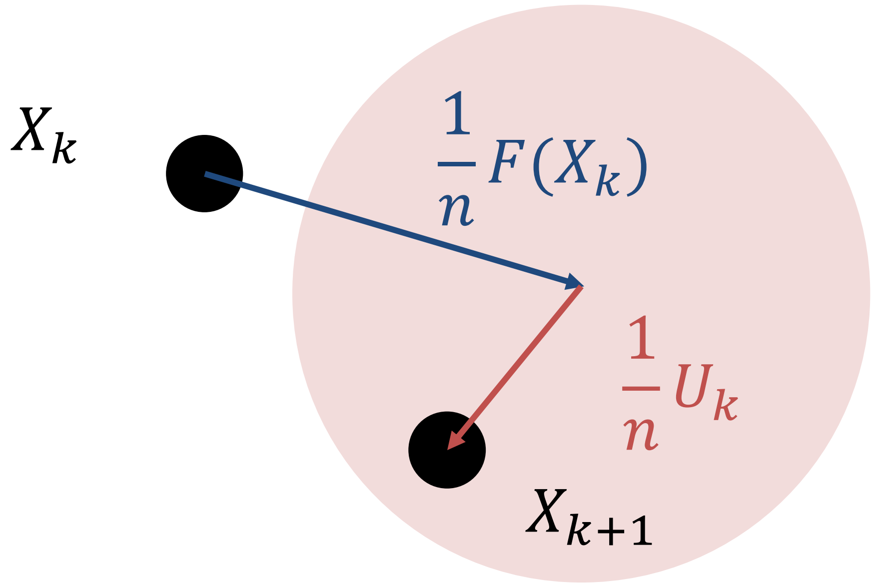


Reinforced random walk with F

A discrete time stochastic process $\{X_k : k = 0, 1, \dots\}$ in $\mathbb{R}^d$ that admits the following representation,

$$X_{k+1} - X_k = \frac{1}{n}\left(F(X_k) + U_k\right)$$

- Expected difference (drift), $F(X)$
- Unbiased noise (noise), U_k

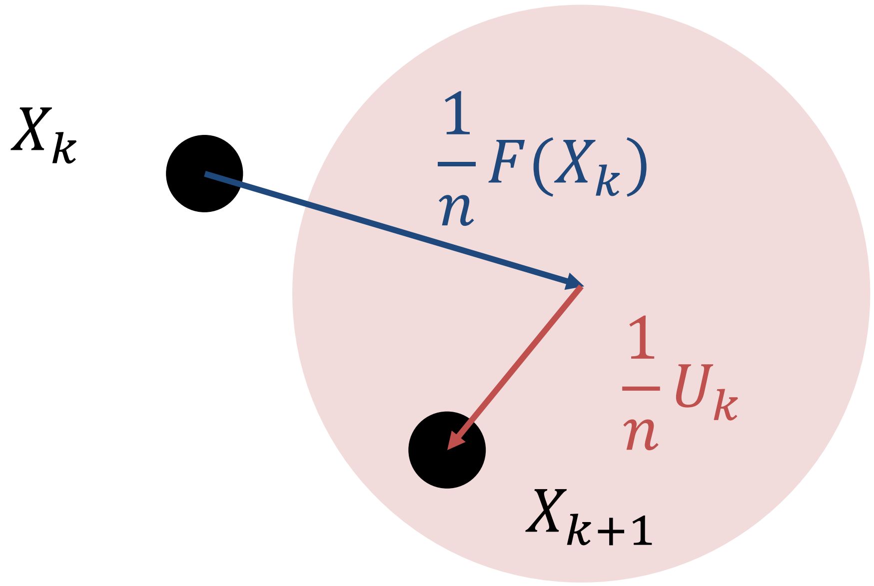


Reinforced random walk with F

A discrete time stochastic process $\{X_k : k = 0, 1, \ldots\}$ in $\mathbb{R}^d$ that admits the following representation,

$$X_{k+1} - X_k = \frac{1}{n}\left(F(X_k) + U_k\right)$$

- Expected difference (drift), $F(X)$
- Unbiased noise (noise), U_k
- Step size, $1/n$



Examples

A discrete time Markov process $\{X_k : k = 0, 1, \dots\}$ in $\mathbb{R}^d$ that admits the following representation,

$$X_{k+1} - X_k = \frac{1}{n}\left(F(X_k) + U_k\right)$$

- Agent based models with n agents
  - Evolutionary games
  - Dynamics on social networks
- Heuristic local search algorithms with uniform step size $1/n$

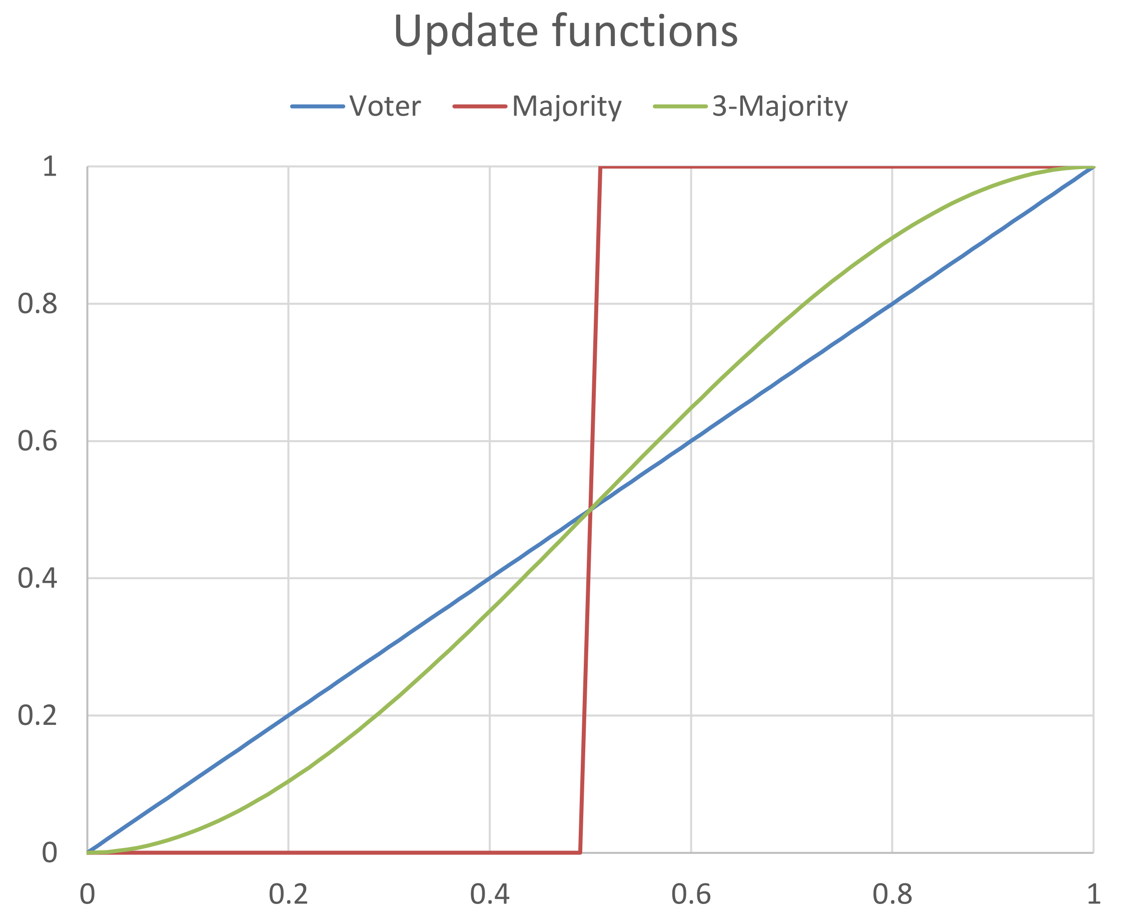
Node Dynamic on complete graphs [SY18]

- Let $f_{ND}: [0,1] \to [0,1]$. n agents interact on a complete graph
- Each agent v has an initial binary state $C_0(v) \in \{0,1\}$
- At round k,
  - Pick a node v uniformly at random
  - Compute the fraction of opinion 1, $X_k = \frac{|C_k^{-1}(1)|}{n}$ <- Complete graph
  - Update $C_{k+1}(v)$ to 1 w.p. $f_{ND}(X_k)$; 0 o.w.

Node Dynamic

Includes several existing dynamics

- Voter model
- Iterative majority [Mossel et al 14]
- Iterative 3-majority [Doerr et al 11]



Node Dynamic

Node dynamic on complete graphs

- Let $f_{ND}: [0,1] \to [0,1]$. There are n agents on a complete graph
- Each agent v has an initial binary state $C_0(v) \in \{0,1\}$
- At round k,
  - Pick a node v uniformly at random
  - Compute the fraction of opinion 1, $X_k = \frac{|C_k^{-1}(1)|}{n}$
  - Update $C_{k+1}(v)$ to 1 w.p. $f_{ND}(X_k)$; 0 o.w.

Reinforced random walk on $\mathbb{R}$

- X_k be the fraction of nodes in state 1 at k.

Node Dynamic

Node dynamic on complete graphs

- Let $f_{ND}: [0,1] \to [0,1]$. There are n agents on a complete graph
- Each agent v has an initial binary state $C_0(v) \in \{0,1\}$
- At round k,
  - Pick a node v uniformly at random
  - Compute the fraction of opinion 1, $X_k = \frac{|C_k^{-1}(1)|}{n}$
  - Update $C_{k+1}(v)$ to 1 w.p. $f_{ND}(X_k)$; 0 o.w.

Reinforced random walk on $\mathbb{R}$

- X_k be the fraction of nodes in state 1 at k.
- Given X_k, the expected number of nodes in state 1 after round k, is $\mathrm{E}[nX_{k+1} \mid X_k] = nX_k + (f_{ND}(X_k) - X_k)$.

Node Dynamic

Node dynamic on complete graphs

- Let $f_{ND}: [0,1] \to [0,1]$. There are n agents on a complete graph
- Each agent v has an initial binary state $C_0(v) \in \{0,1\}$
- At round k,
  - Pick a node v uniformly at random
  - Compute the fraction of opinion 1, $X_k = \frac{|C_k^{-1}(1)|}{n}$
  - Update $C_{k+1}(v)$ to 1 w.p. $f_{ND}(X_k)$; 0 o.w.

Reinforced random walk on $\mathbb{R}$

- X_k be the fraction of nodes in state 1 at k.
- Given X_k, the expected number of nodes in state 1 after round k, is $\mathrm{E}[nX_{k+1} \mid X_k] = nX_k + (f_{ND}(X_k) - X_k)$.

Updated to 1 from 1

Node Dynamic

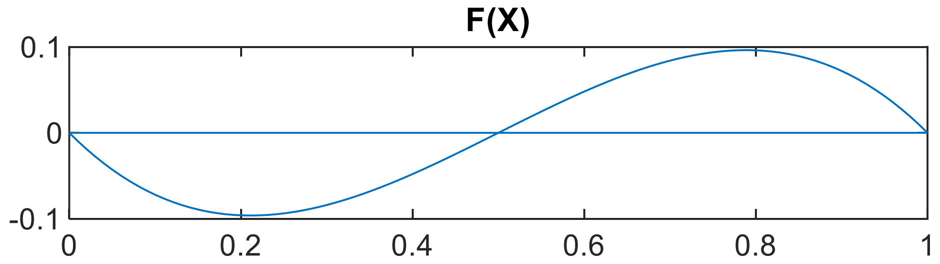
Node dynamic on complete graphs

- Let $f_{ND}: [0,1] \to [0,1]$. There are n agents on a complete graph
- Each agent v has an initial binary state $C_0(v) \in \{0,1\}$
- At round k,
  - Pick a node v uniformly at random
  - Compute the fraction of opinion 1, $X_k = \frac{|C_k^{-1}(1)|}{n}$
  - Update $C_{k+1}(v)$ to 1 w.p. $f_{ND}(X_k)$; 0 o.w.

Reinforced random walk on $\mathbb{R}$

- X_k be the fraction of nodes in state 1 at k.
- $\mathrm{E}[X_{k+1} \mid X_k] - X_k = \frac{1}{n}(f_{ND}(X_k) - X_k)$.

Drift $F(X_k)$



Node Dynamic

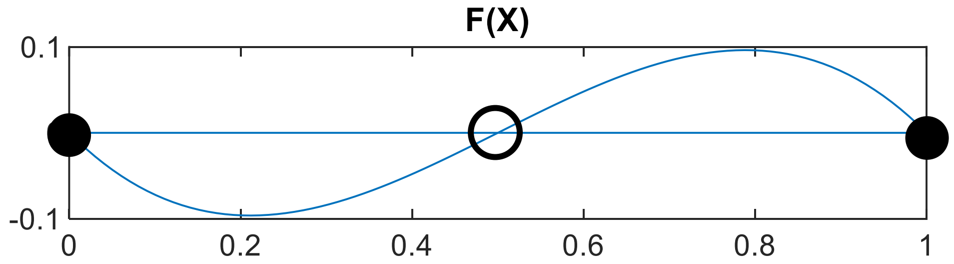
Node dynamic on complete graphs

- Let $f_{ND}: [0,1] \to [0,1]$. There are n agents on a complete graph
- Each agent v has an initial binary state $C_0(v) \in \{0,1\}$
- At round k,
  - Pick a node v uniformly at random
  - Compute the fraction of opinion 1, $X_k = \frac{|C_k^{-1}(1)|}{n}$
  - Update $C_{k+1}(v)$ to 1 w.p. $f_{ND}(X_k)$; 0 o.w.

Reinforced random walk on $\mathbb{R}$

- X_k be the fraction of nodes in state 1 at k.
- $X_{k+1} - X_k = \frac{1}{n}\left((f_{ND}(X_k) - X_k) + U_k\right).$

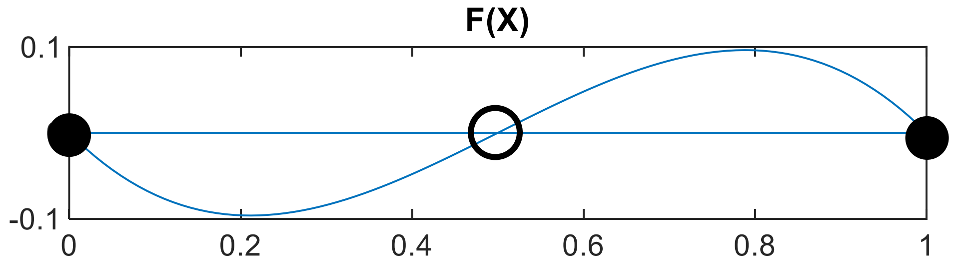
Drift Noise



Question

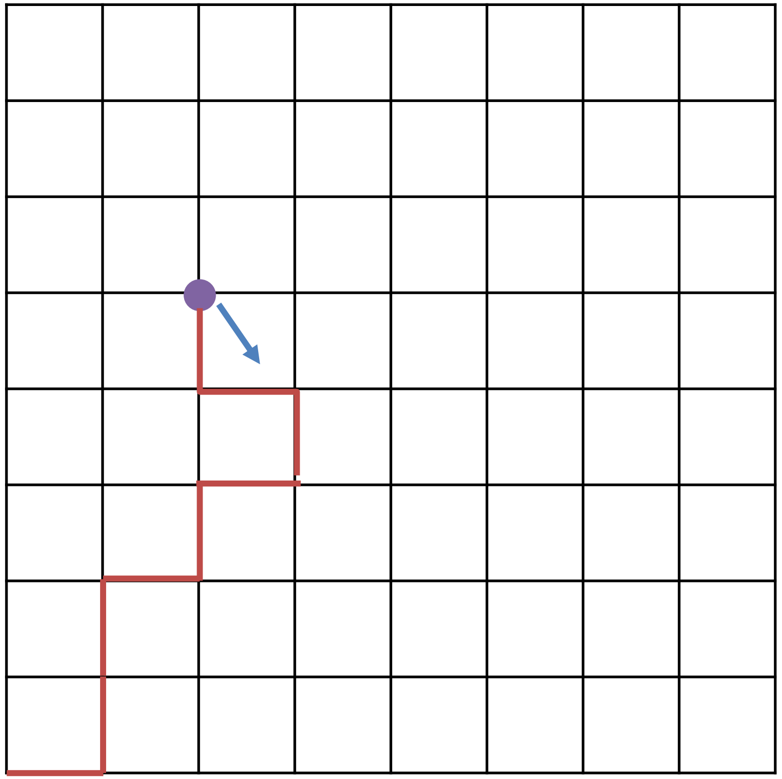
Given F and U, what is the limit of X_k for sufficiently large n?

$$X_{k+1} - X_k = \frac{1}{n}\left(F(X_k) + U_k\right)$$

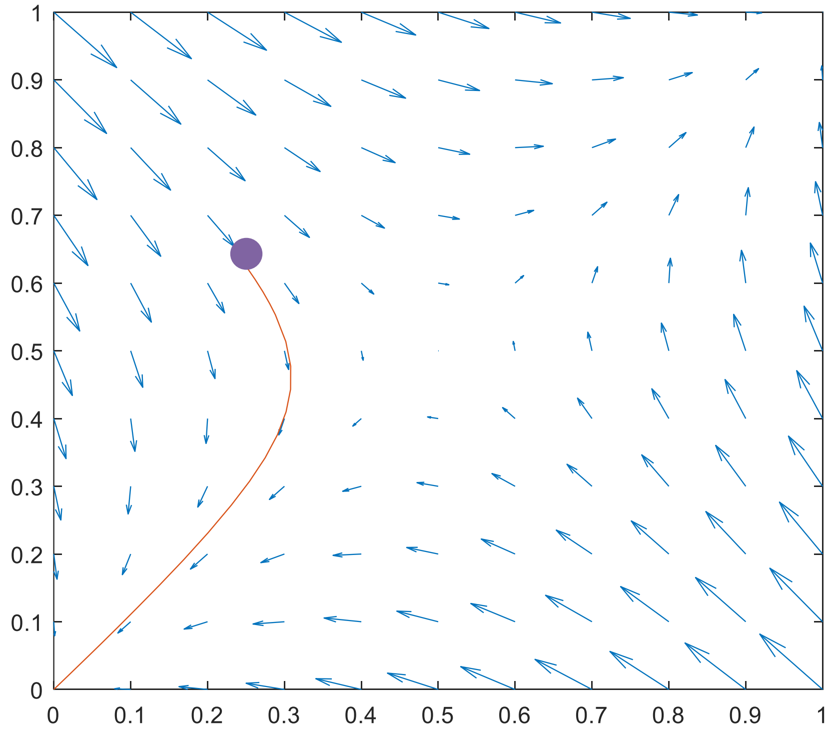


Mean field approximation

$$X_{k+1} - X_k = \frac{1}{n}(F(X_k) + U(X_k))$$



$$x' = F(x)$$

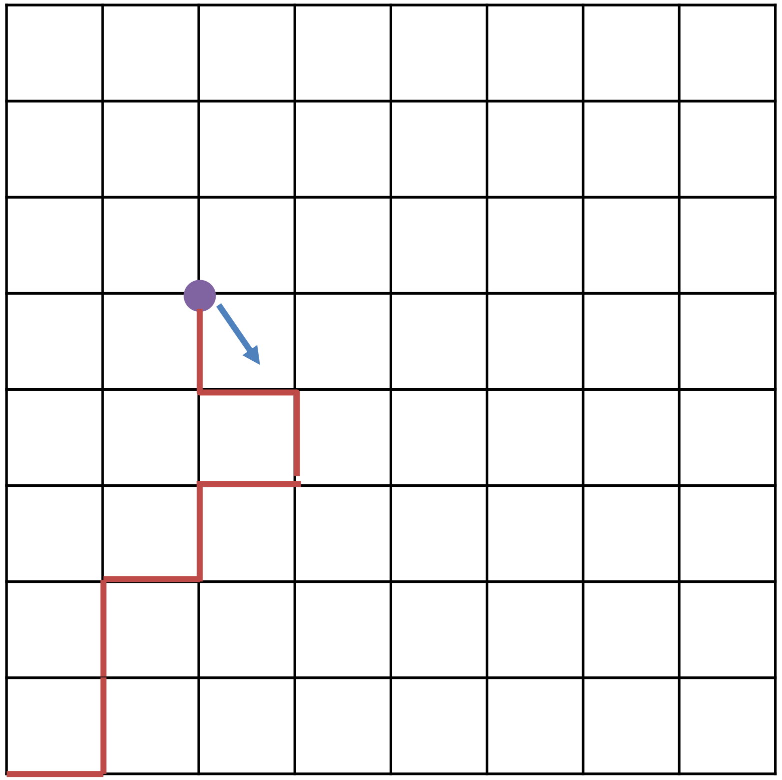


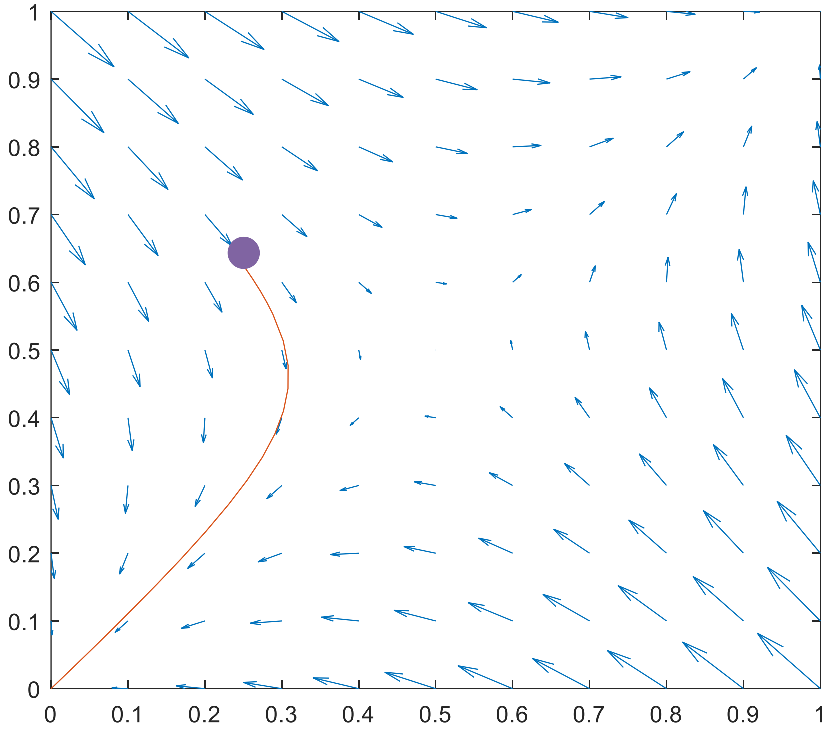
Mean field approximation

If n is large enough, for $k = O(n)$, $X_k \approx x\left(\frac{k}{n}\right)$ by Wormald et al 95.

Regular point

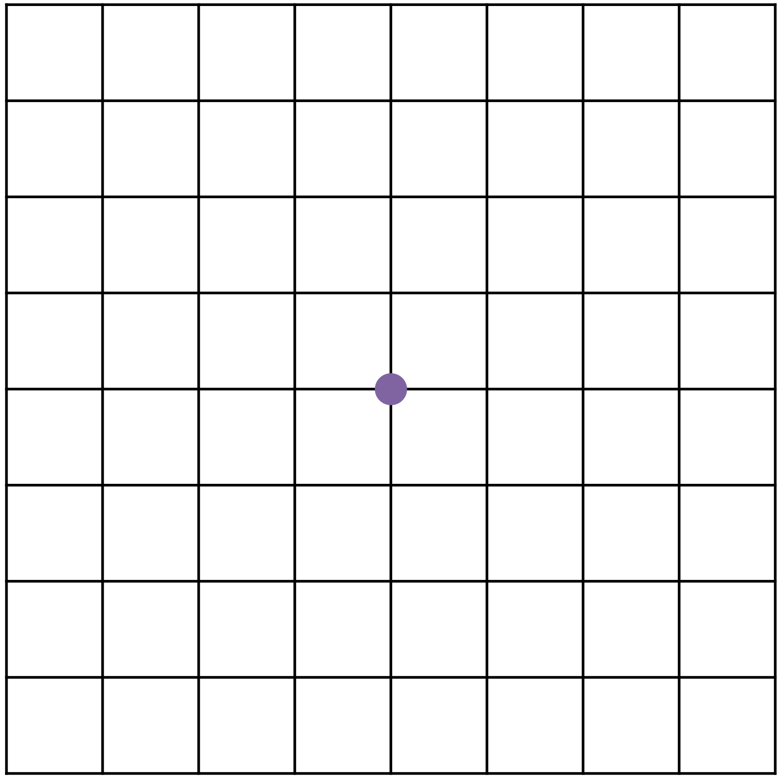
If n is large enough, for $k = O(n)$, $X_k \approx x\left(\frac{k}{n}\right)$.

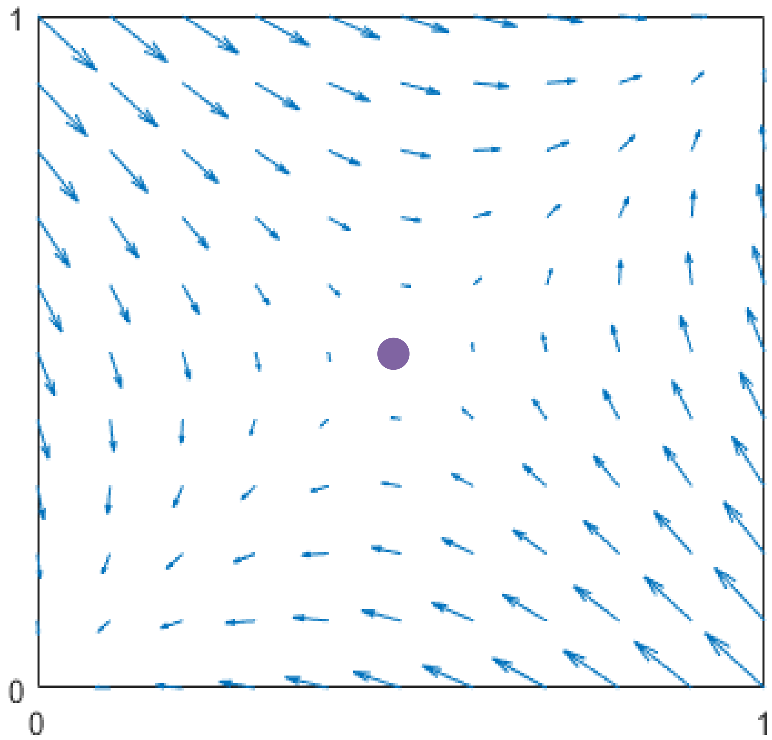




Fixed point, $\boldsymbol{F}(\boldsymbol{x}^*) = \mathbf{0}$

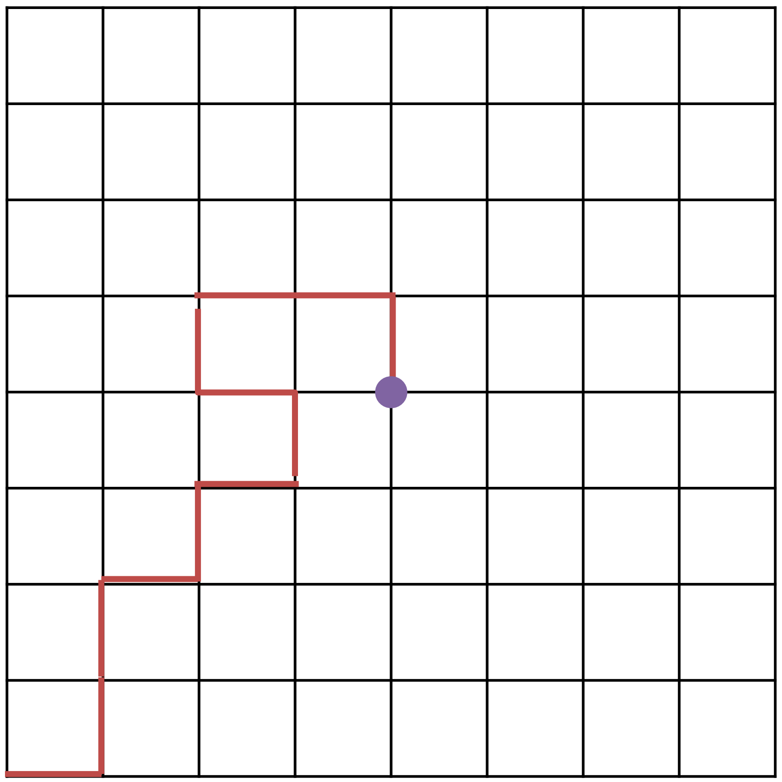
If n is large enough, for $k = O(n)$, $X_k \approx x\left(\frac{k}{n}\right)$.

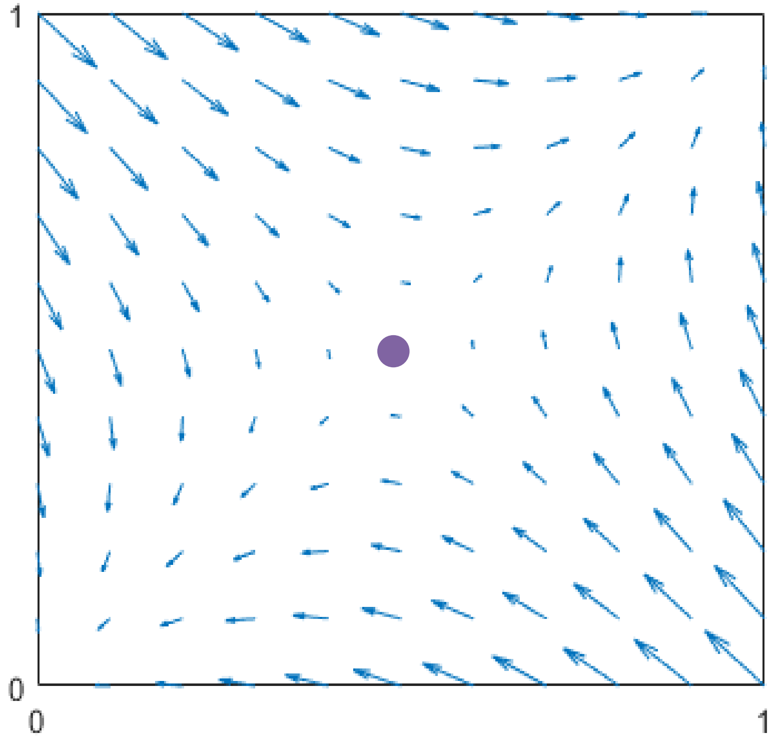




Escaping non-attracting fixed point

When can the process escape a non-attracting fixed point?

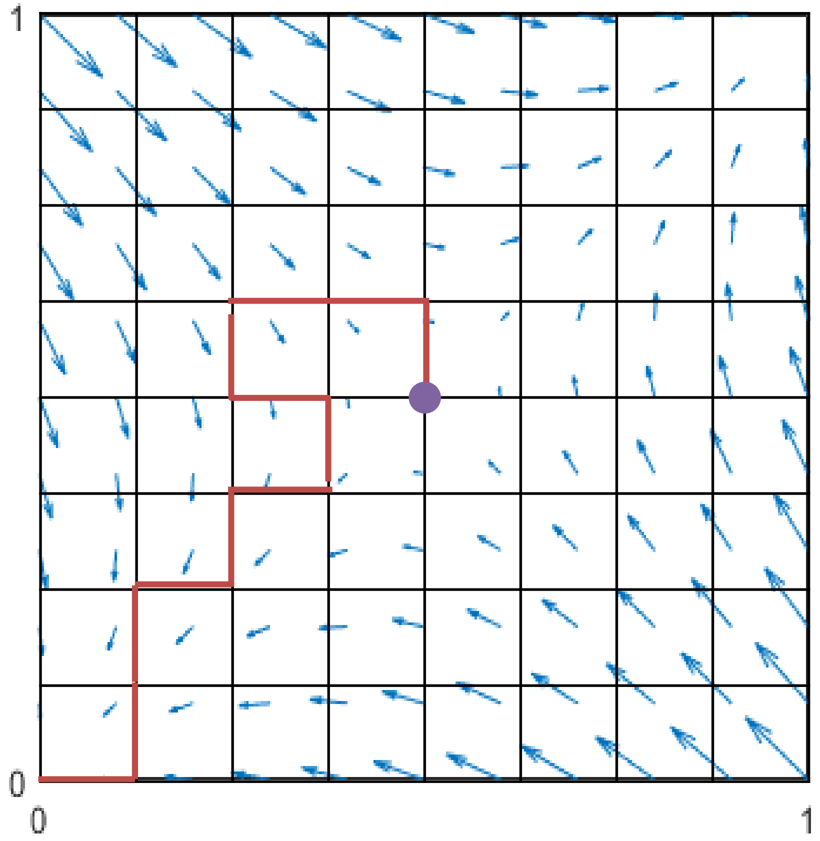




Escaping non-attracting fixed point

When can the process escape a non-attracting fixed point?

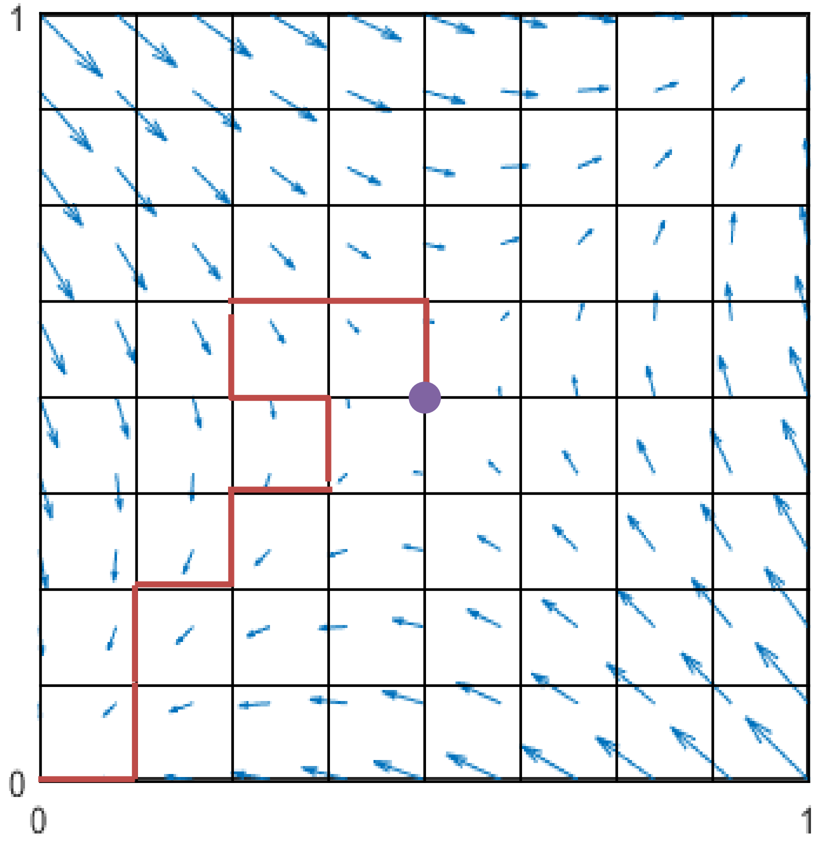
1. $\Theta(n)$
2. $\Theta(n \log n)$
3. $\Theta(n\ (\log n)^4)$
4. $\Theta(n^2)$



Escaping non-attracting fixed point

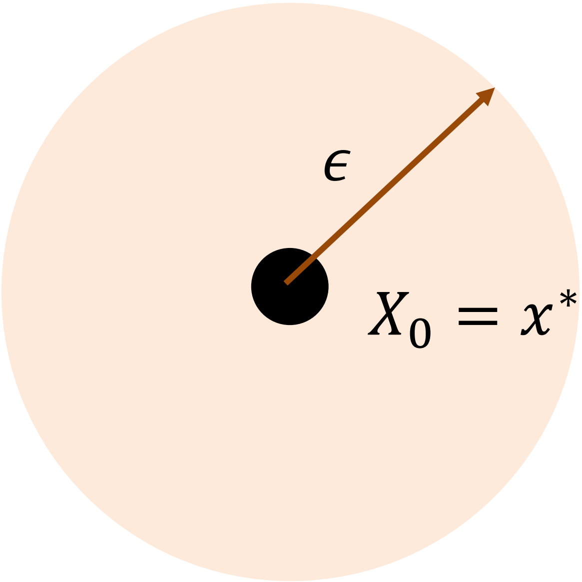
When can the process escape a non-attracting fixed point?

1. $\Theta(n)$
2. $\boldsymbol{\Theta(n\, log\, n)}$
3. $\Theta(n\,(\log n)^4)$
4. $\Theta(n^2)$



Lower bound

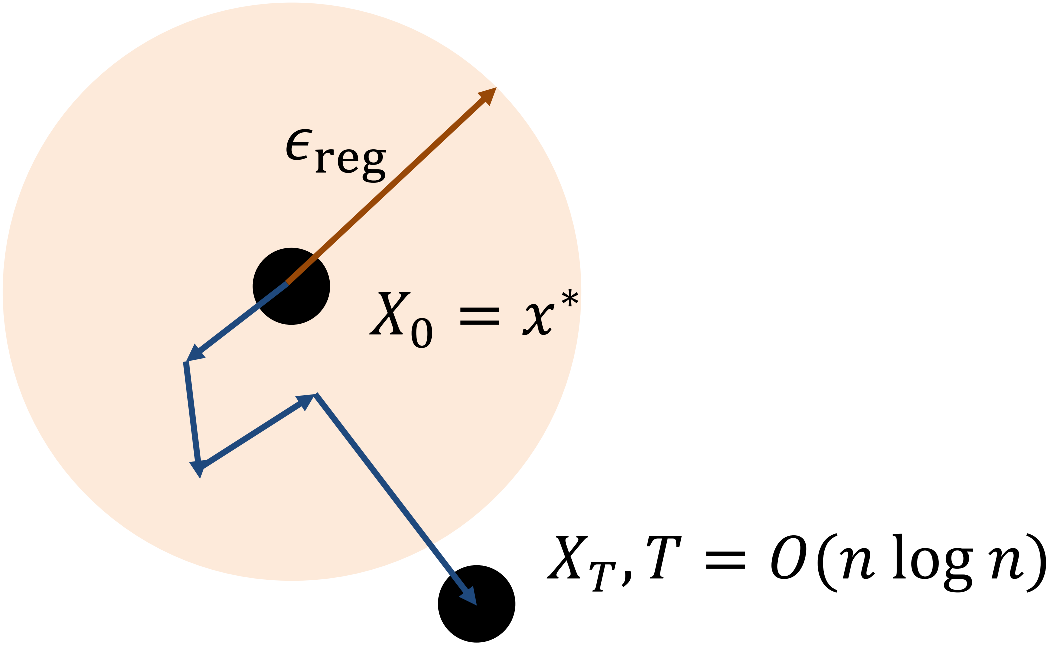
Escaping saddle point region takes at least $\Omega(n \log n)$ steps.



Upper bound

Escaping saddle point region takes at most $O(n \log n)$ steps.

If

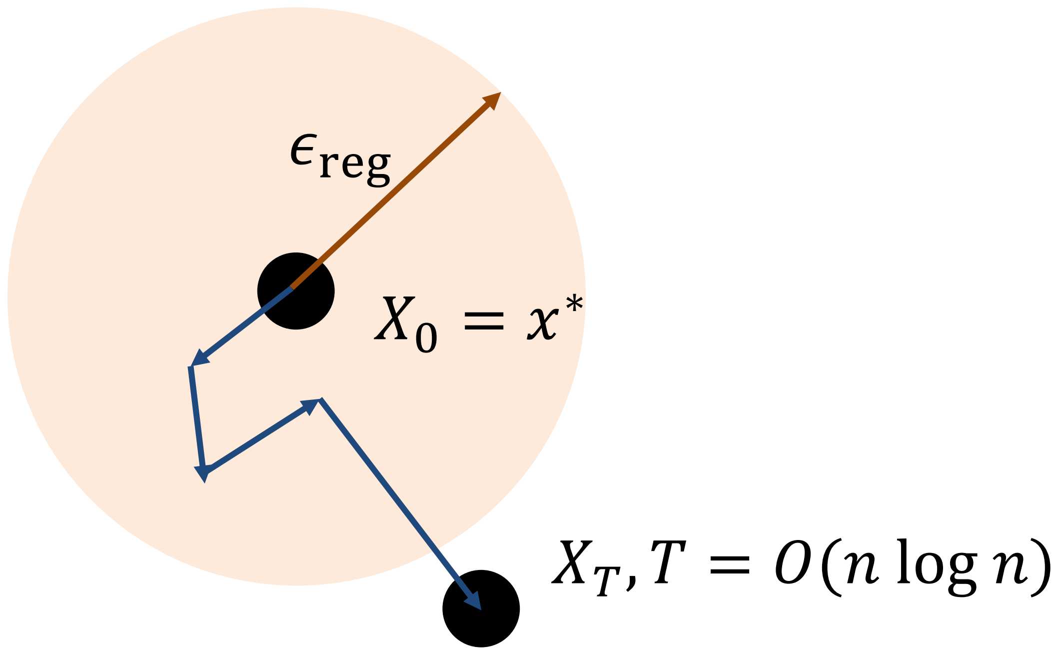


Upper bound

Escaping saddle point region takes at most $O(n \log n)$ steps.

If

- Noise, U_k
  - Martingale difference
  - bounded
  - Noisy (covariance matrix is large)
- Expected difference, $F \in \mathcal{C}^2$
  - x^* is hyperbolic

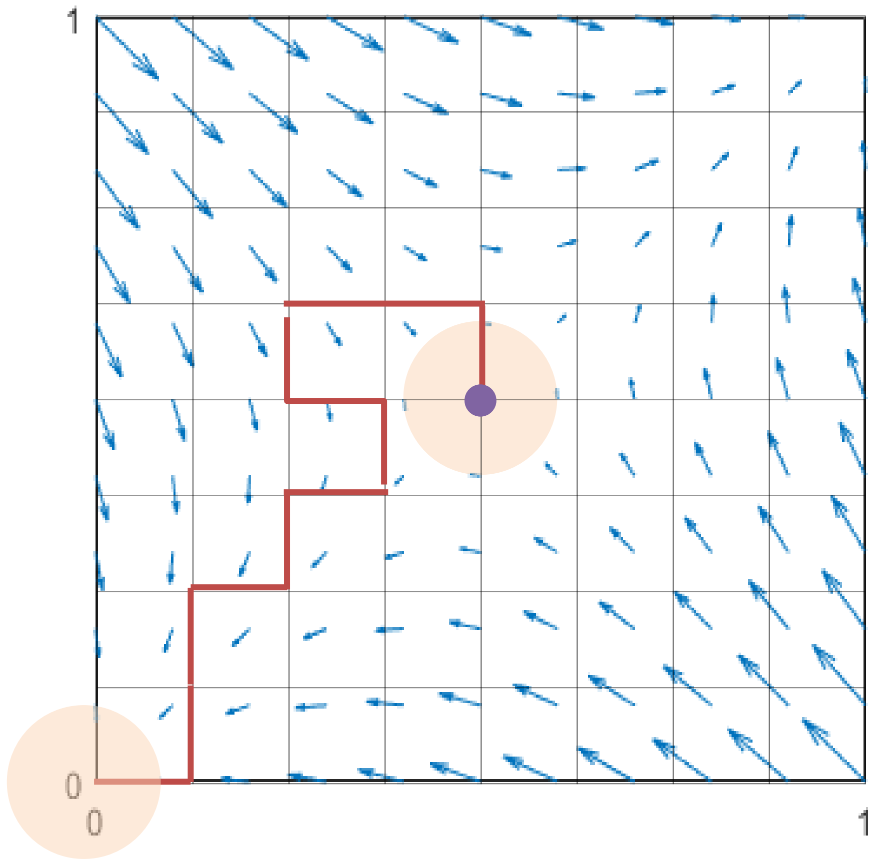


Gradient-like dynamics

Converges to an attracting fixed-point region in $O(n \log n)$ steps.

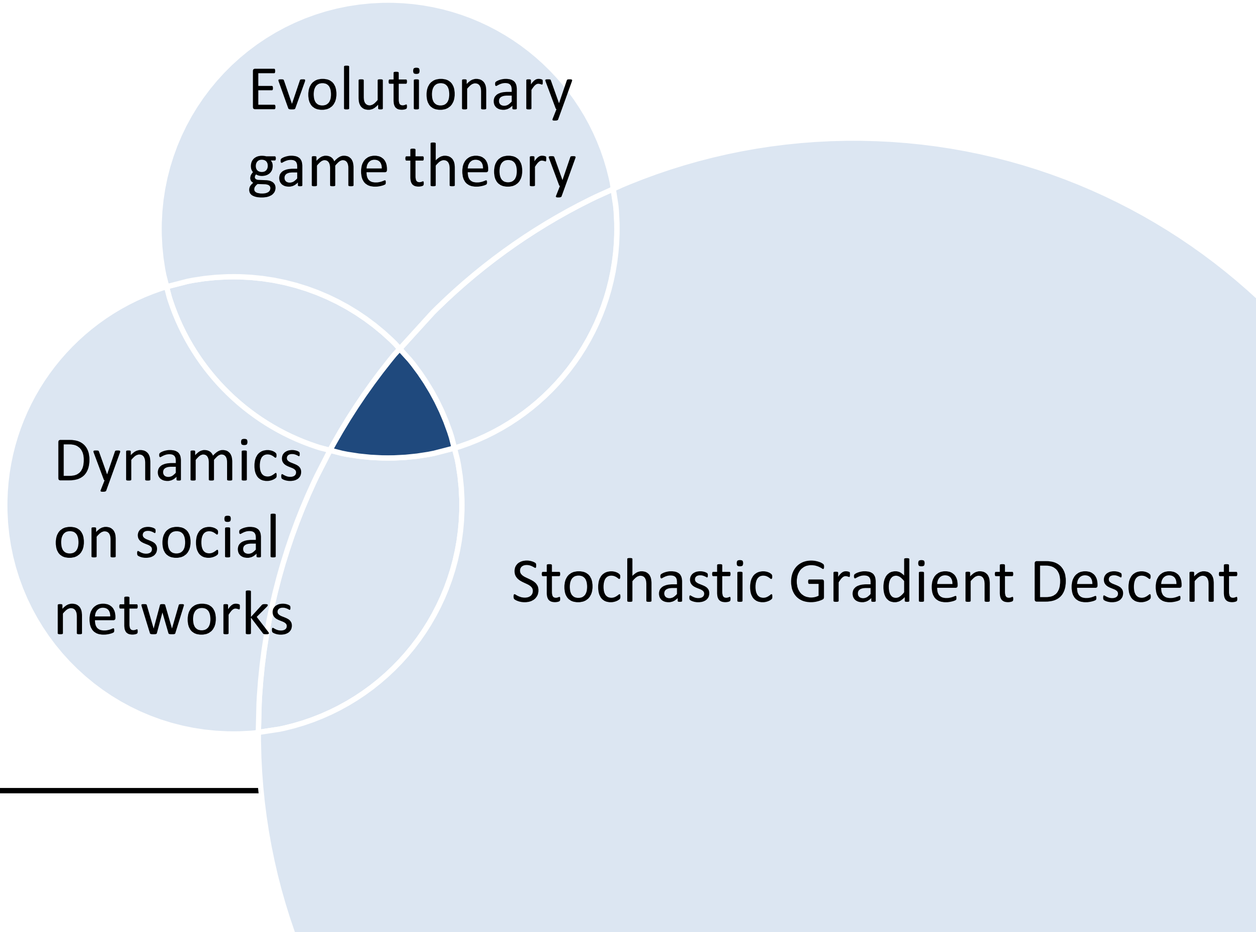
If

- Noise, U_k
  - Martingale difference
  - bounded
  - Noisy
- Expected difference, $F \in \mathcal{C}^2$
  - Fixed points are hyperbolic
  - Potential function



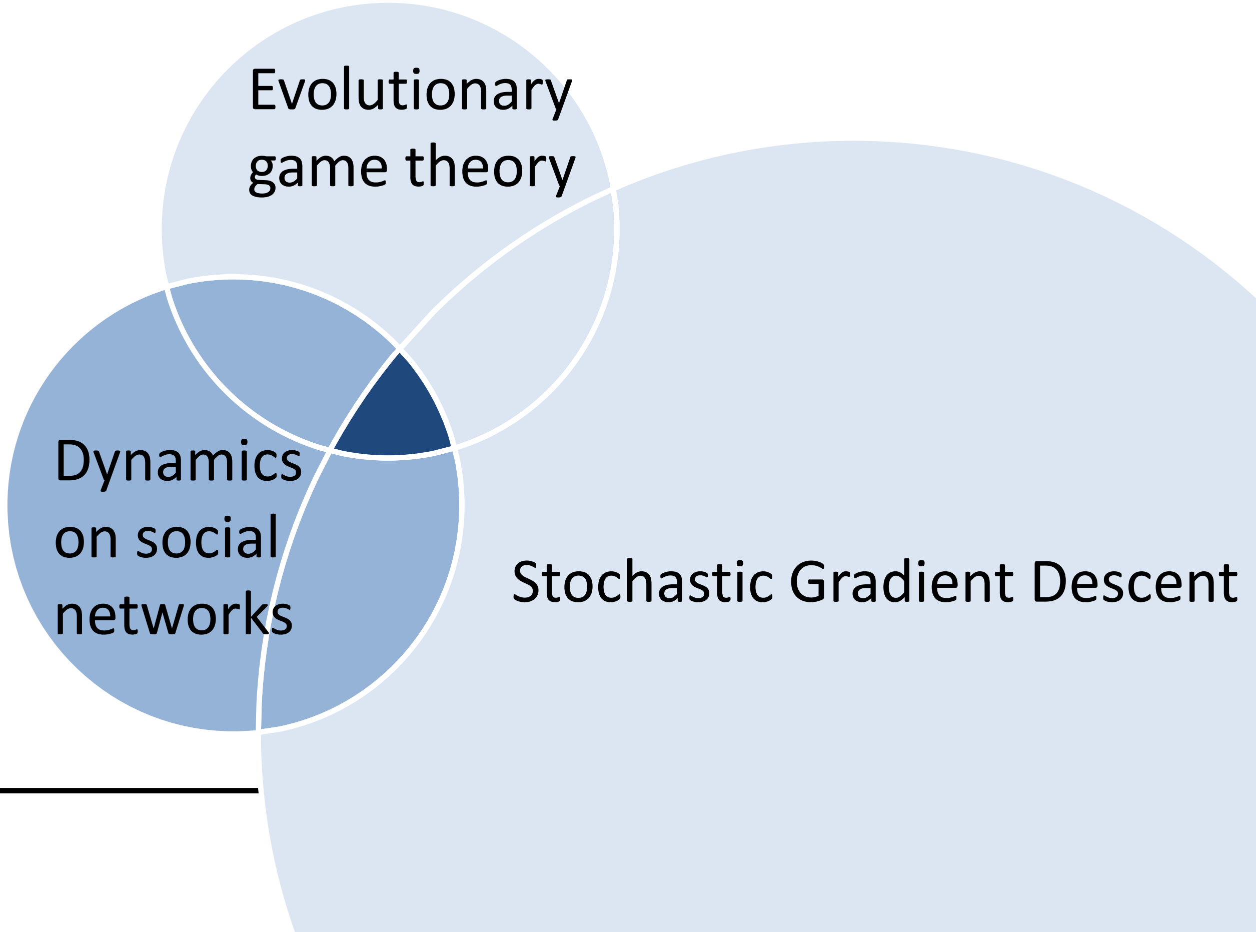
Outline

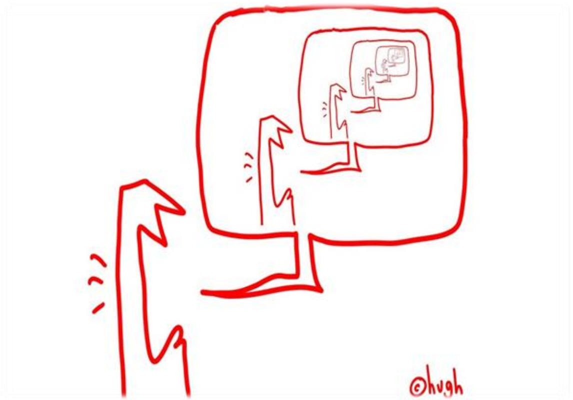
- Escaping saddle point



Outline

- Escaping saddle point
- Case study: dynamics on social networks



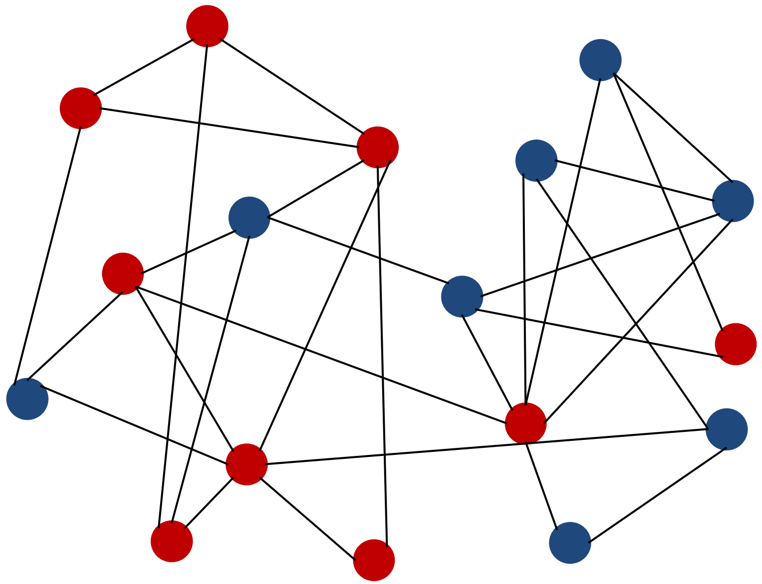


Dynamics on social networks

(DIS)AGREEMENT IN PLANTED COMMUNITY NETWORKS

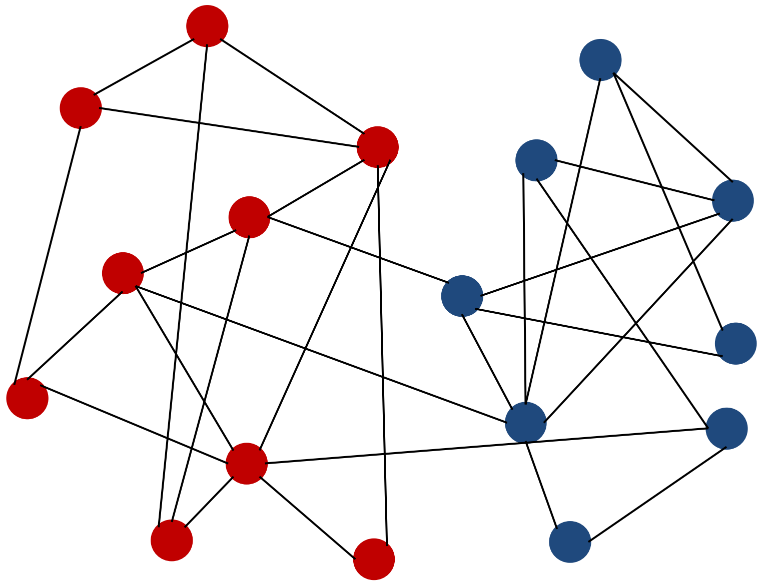
Echo chamber

Beliefs are amplified through interactions in segregated systems



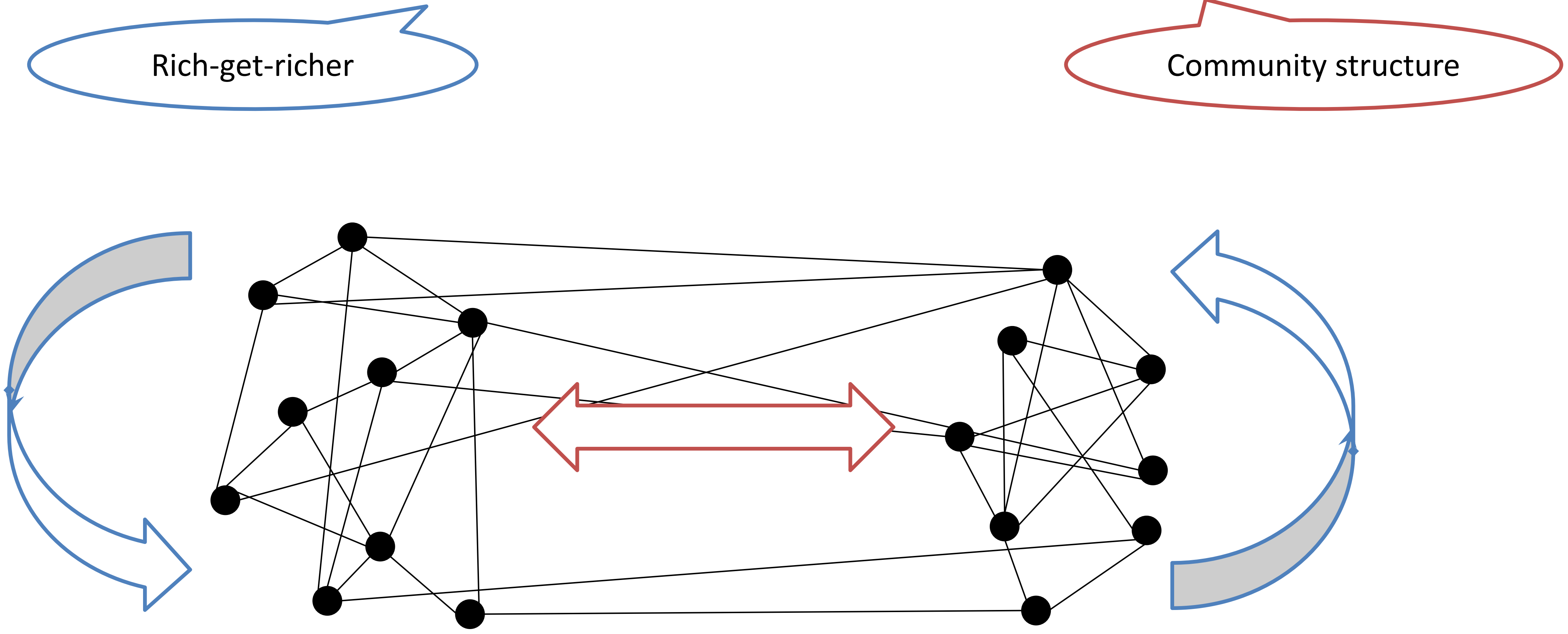
Echo chamber

Beliefs are amplified through interactions in segregated systems



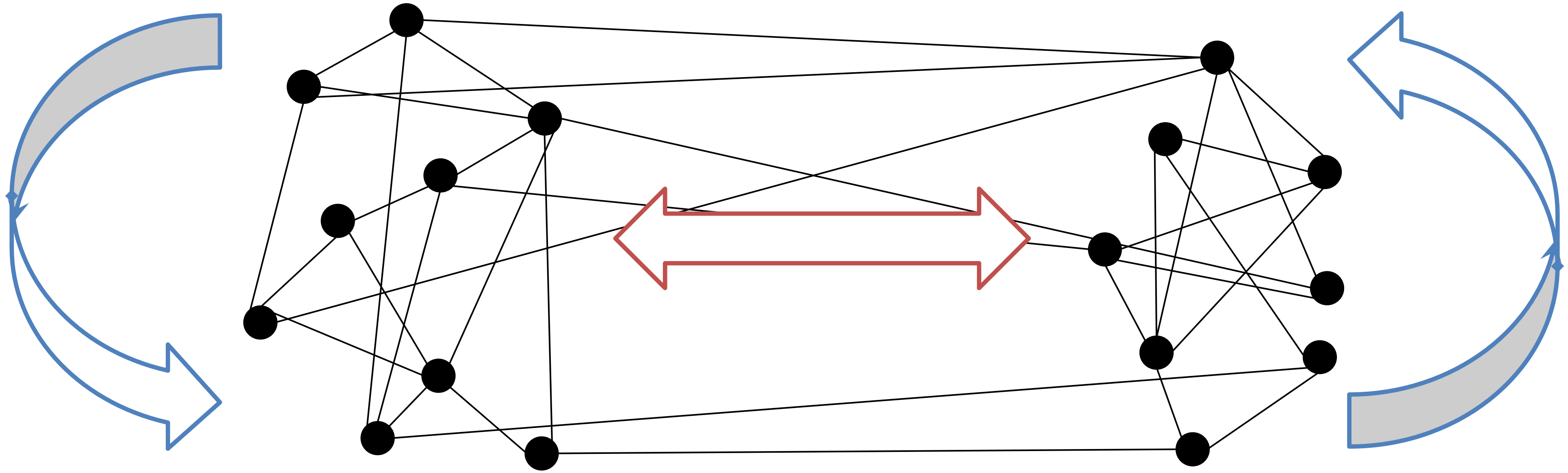
Echo chamber

Beliefs are amplified through interactions in segregated systems



Question

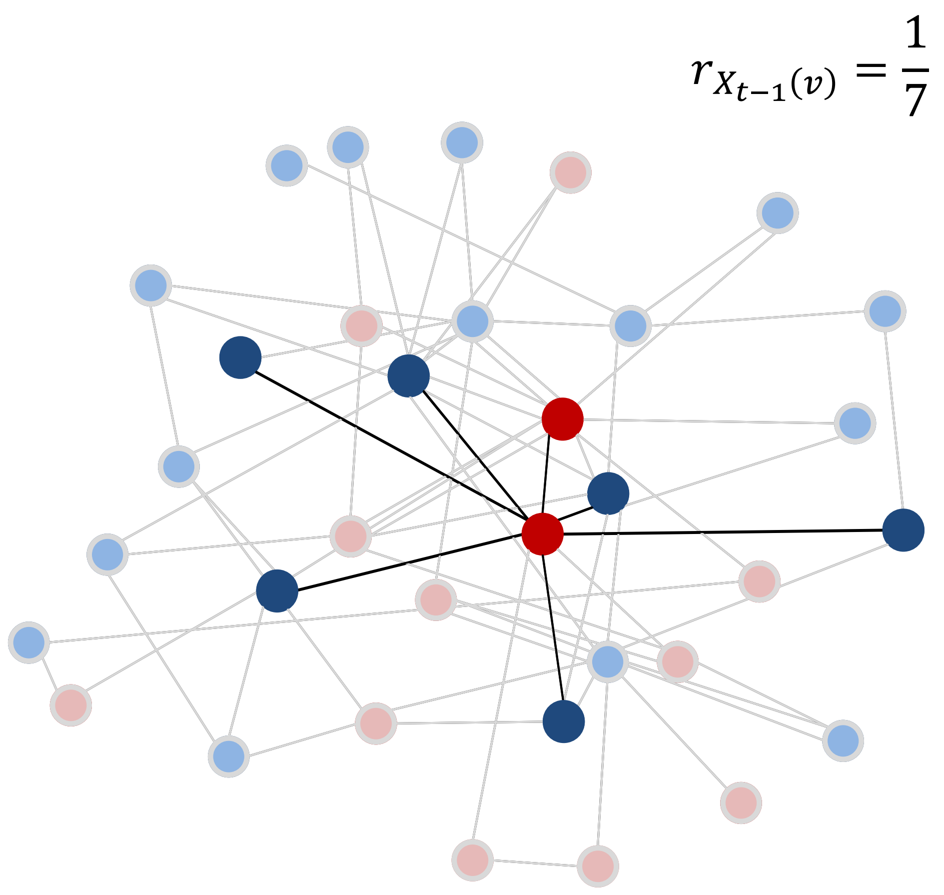
What is the consensus time given a rich-get-richer opinion formation and the level of intercommunity connectivity?



Node Dynamic [Schoenebeck, Yu 18]

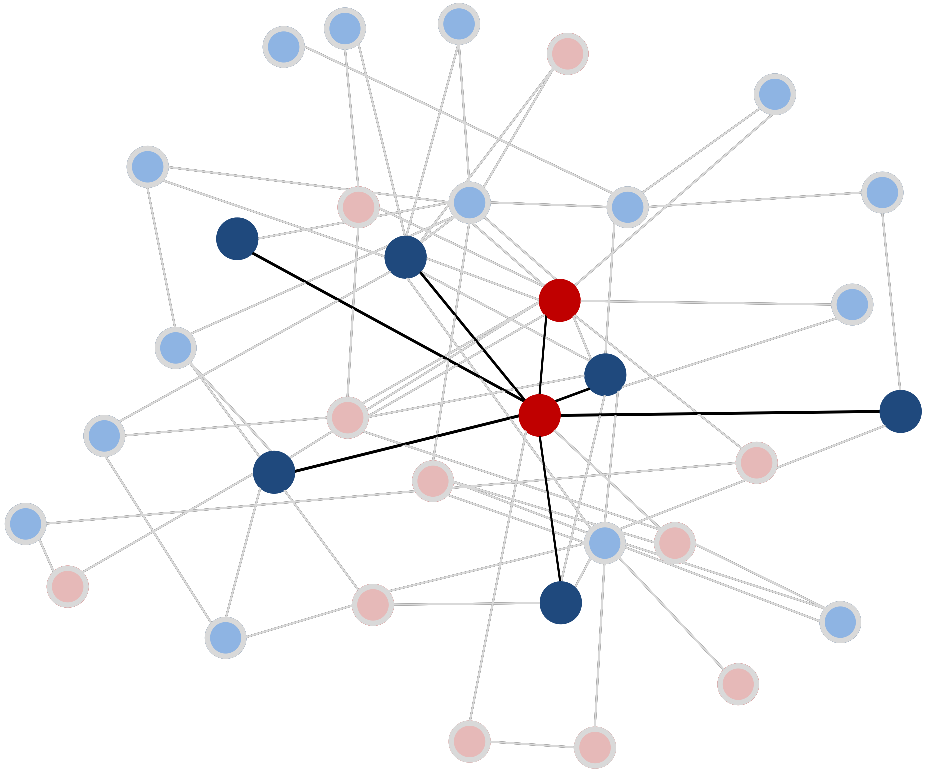
- Fixed a graph $G = (V, E)$ opinion set $\{0,1\}$
- Given an initial configuration $X_0 : V \mapsto \{0,1\}$
- At round t,
  - A node v is picked uniformly at random

The update of opinion only depends on the fraction of opinions amongst its neighbors



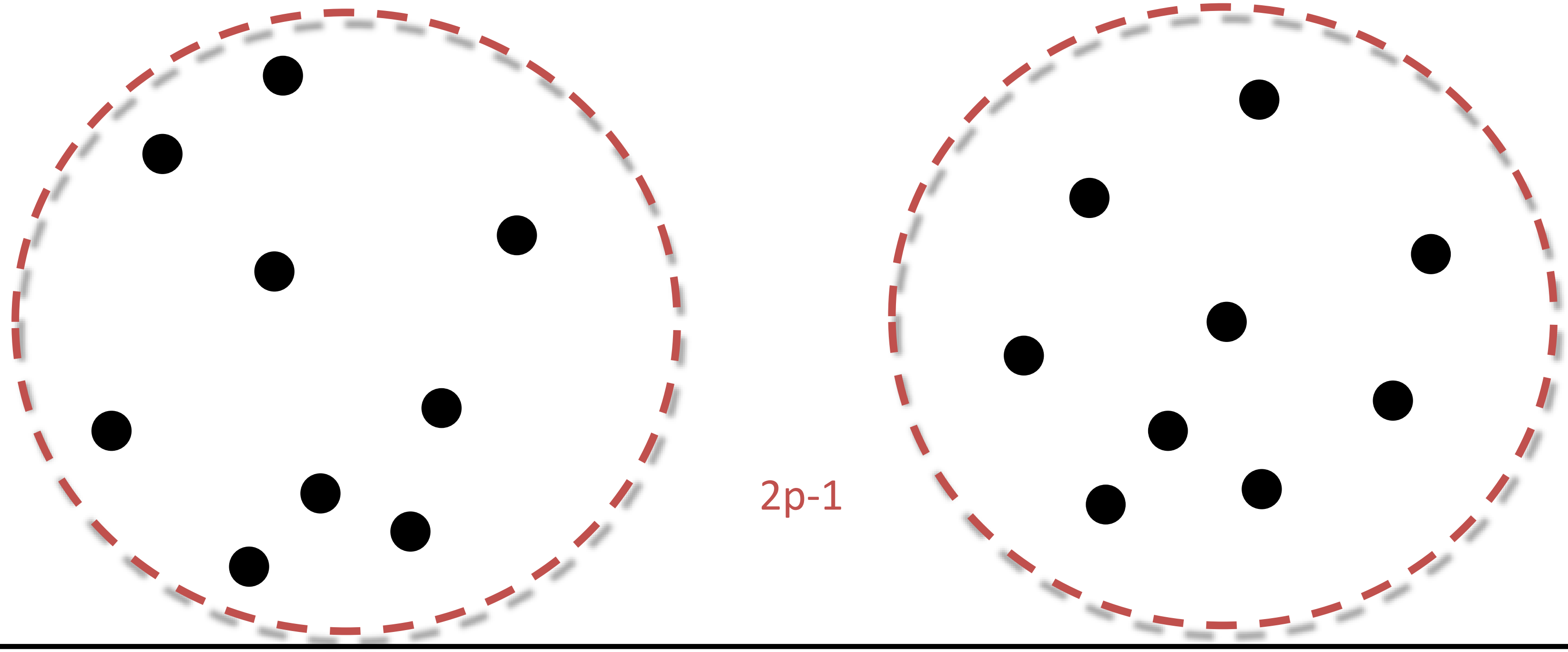
Node Dynamic ND(G, f_{ND}, X_0)

- Fixed a (weighted) graph $G = (V, E)$ opinion set $\{0,1\}$, **an update function** $\boldsymbol{f_{ND}}$
- Given an initial configuration $X_0 : V \mapsto \{0,1\}$
- At round t,
  - A node v is picked uniformly at random
  - $\boldsymbol{X_t(v) = 1}$ ***w.p.*** $\boldsymbol{f_{ND}\left(r_{X_{t-1}(v)}\right)}$ **;**
    $= 0$ **otherwise**



Planted Community

- A weighted complete graph with n nodes, $K(n, p)$
  - Two communities with equal size
  - An edge has weight p if in the same community and $1 - p$ o.w.

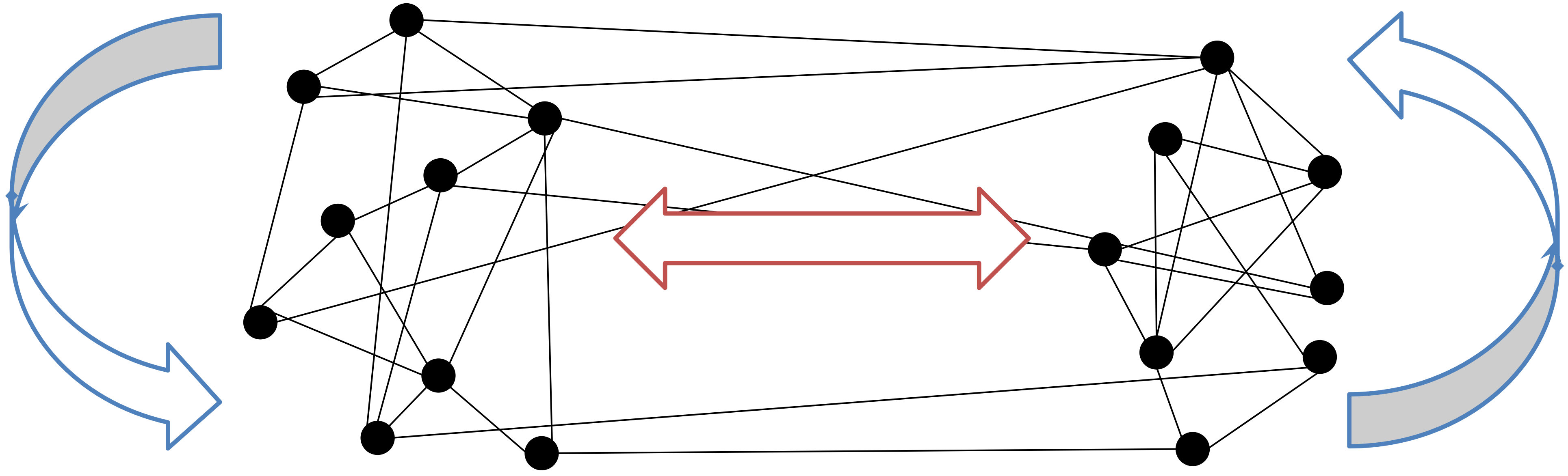


Planted Community

- A weighted complete graph with n nodes, $K(n, p)$
  - Two communities with equal size
  - An edge has weight p if in the same community and $1 - p$ o.w.

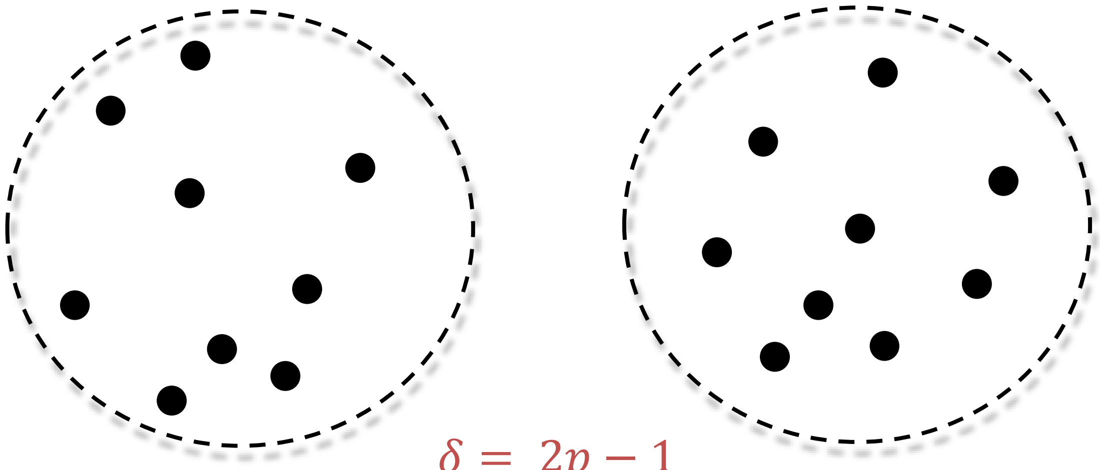
Question

- What is the interaction between rich-get-richer opinion formation and the level of intercommunity connectivity?

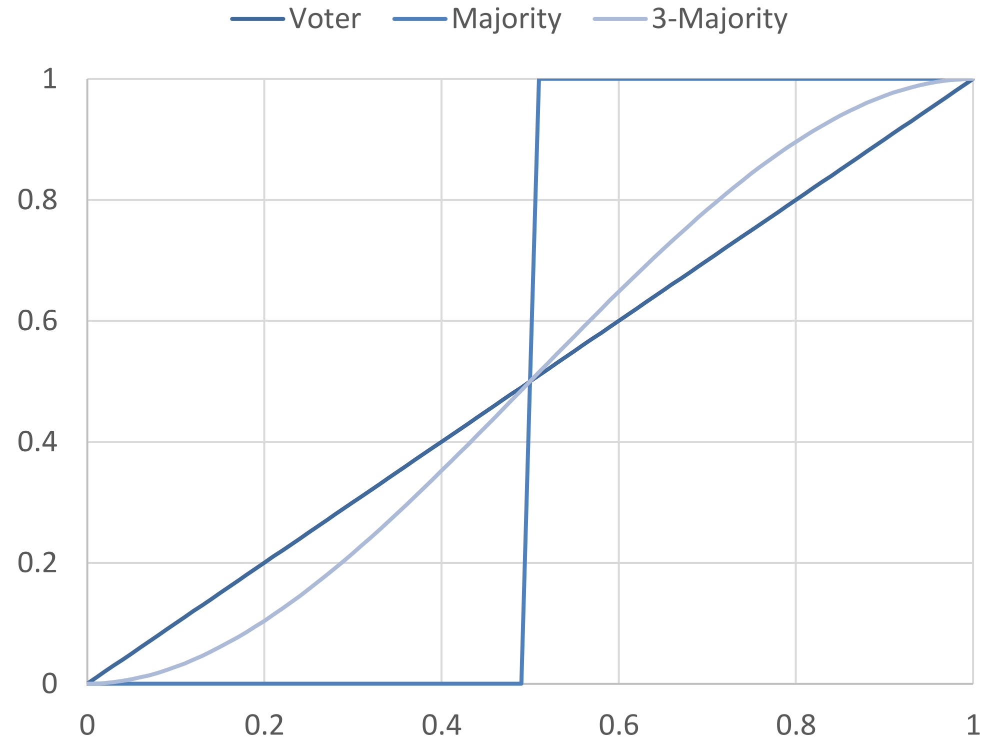


Question

- What is the interaction between rich-get-richer opinion formation and the level of intercommunity connectivity?

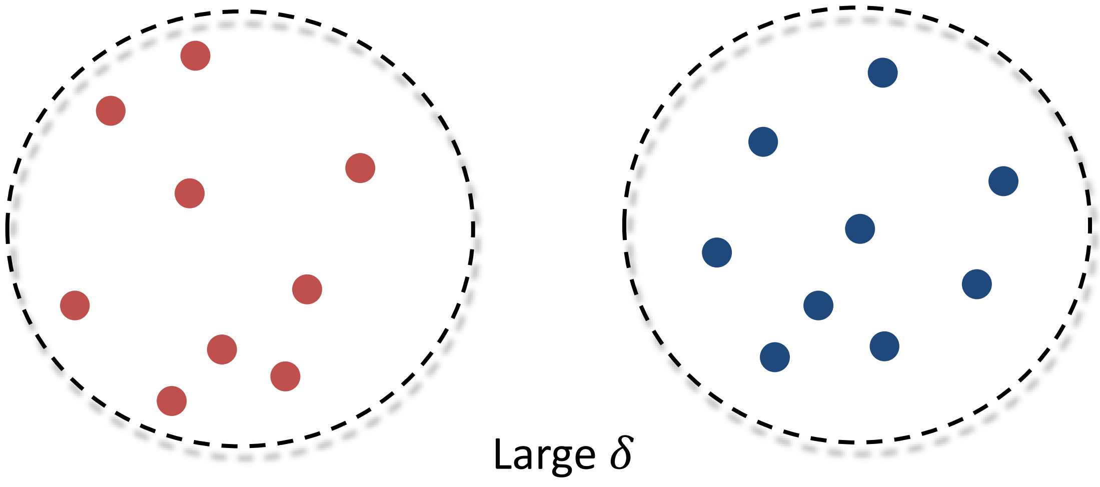


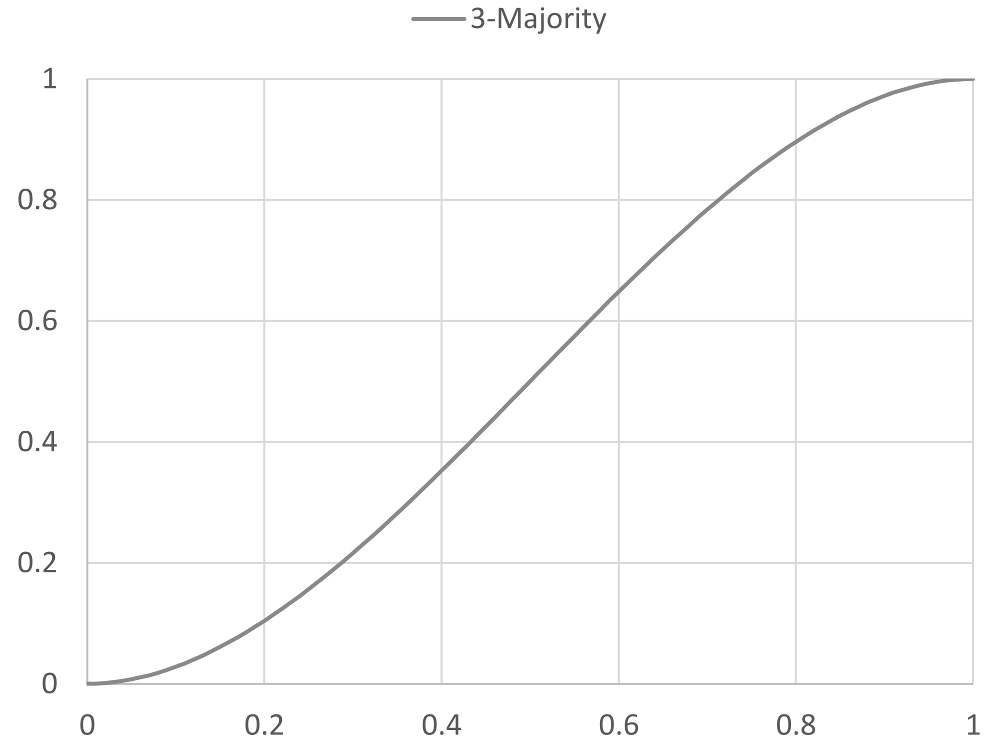
$\delta = 2p - 1$



Strong Community Structure

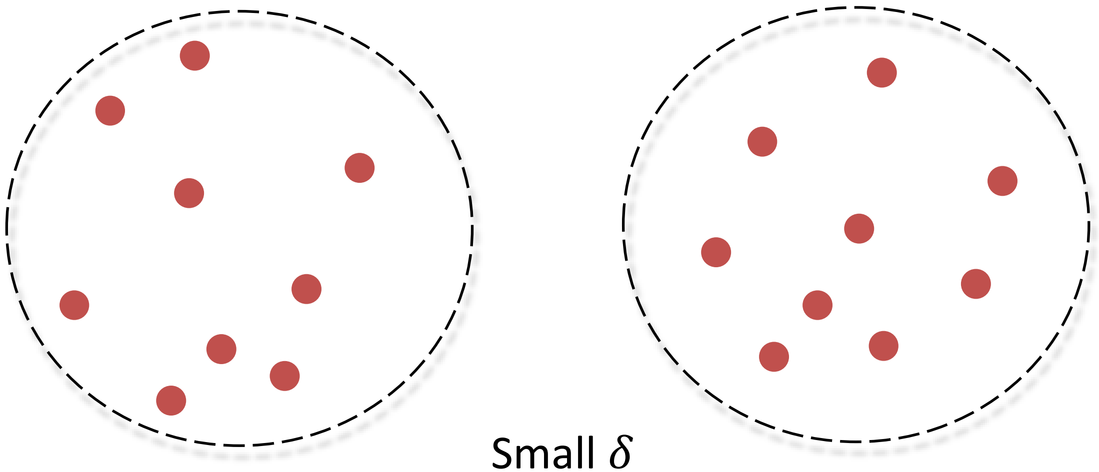
- There exists an initial state such that the process cannot reach consensus fast.



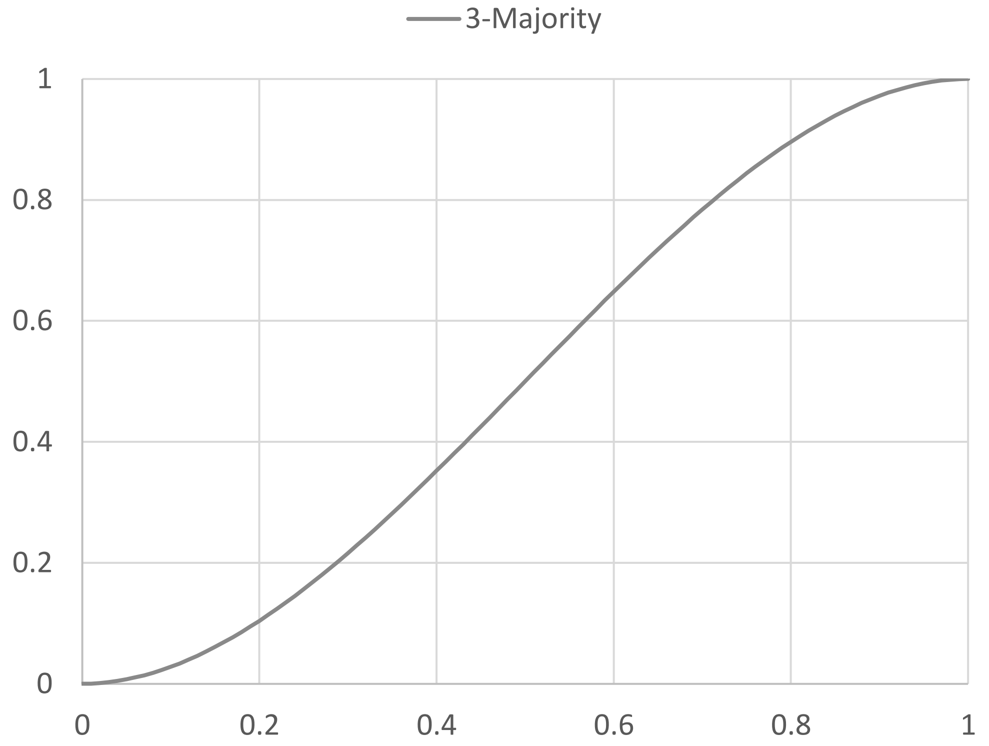


Weak Community Structure

- For all initial states, the process reaches consensus fast.

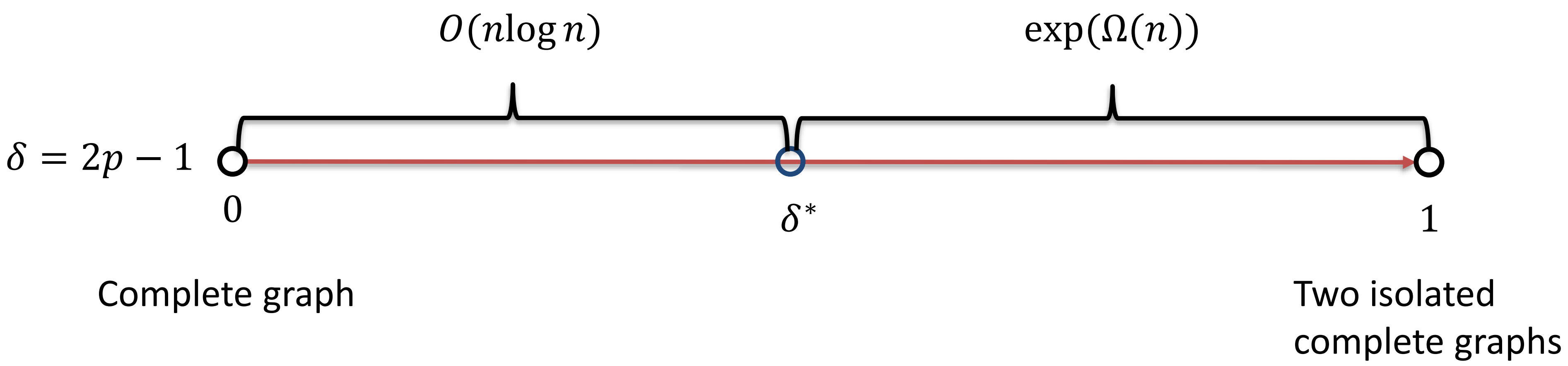


Small δ



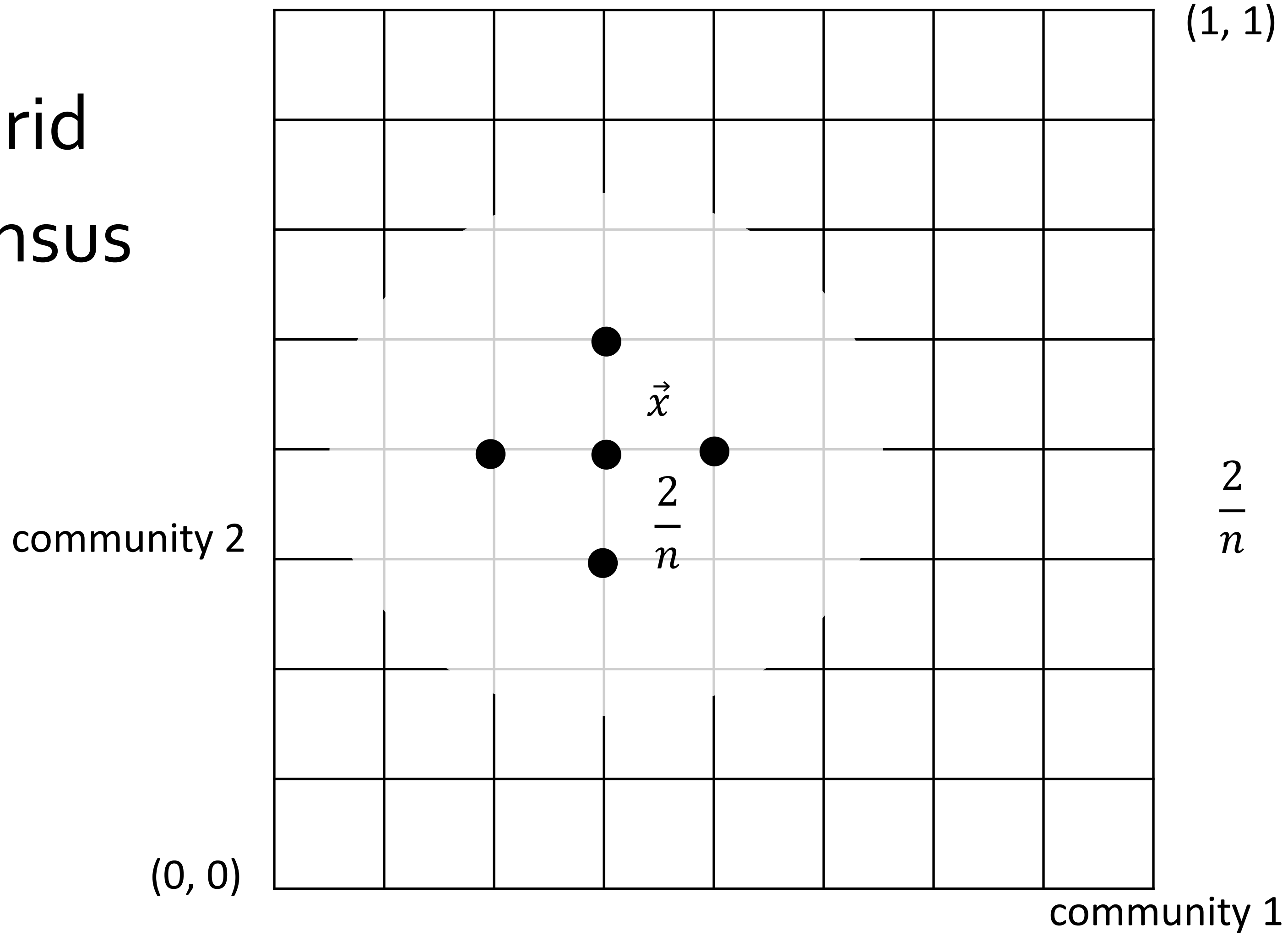
Our Dichotomy Theorem

- Given a smooth rich-get-richer function $f_{ND} \in \mathcal{C}^2$, and a planted community graph $G = K(n, p)$. The maximum expected consensus time of $\mathrm{ND}(G, f_{ND}, X_{\mathbf{0}})$ has two cases:

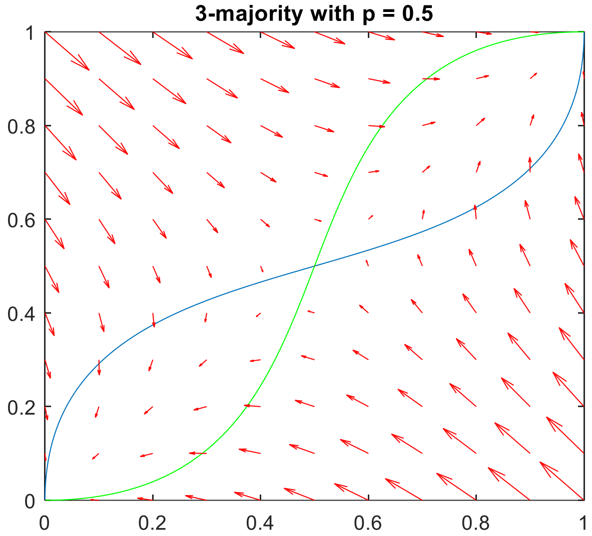


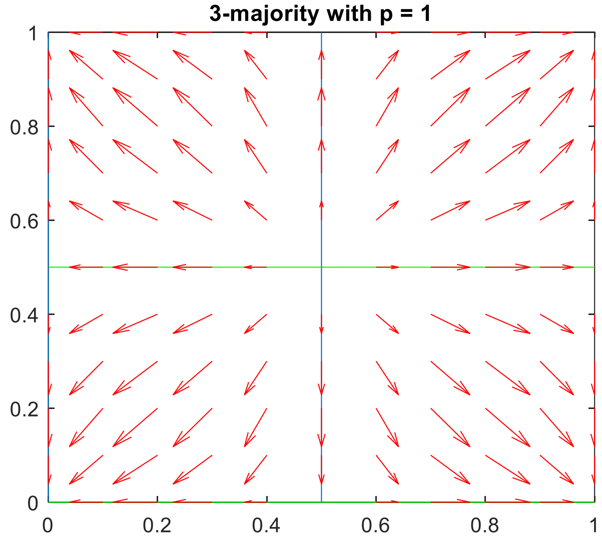
Node dynamic

- A Markov chain on 2-d grid
- (0,0) and (1,1) are consensus states



Our Dichotomy Theorem





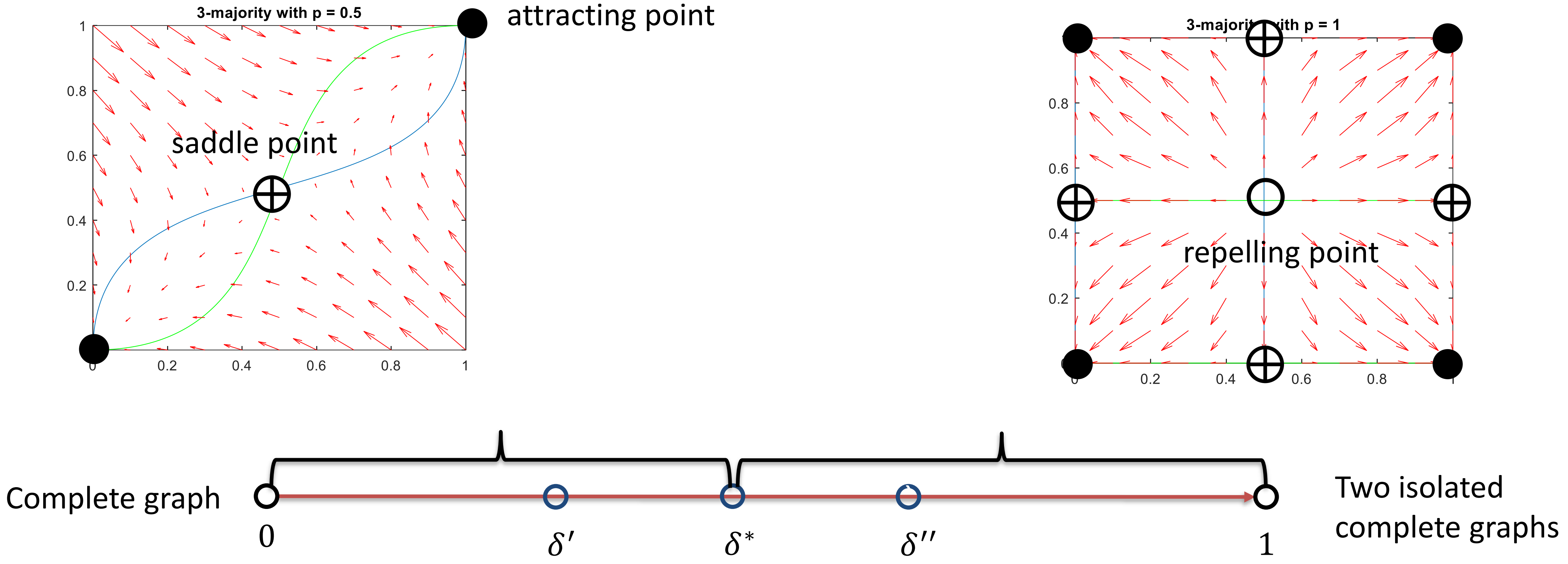
Complete graph

$\delta = 2p - 1$

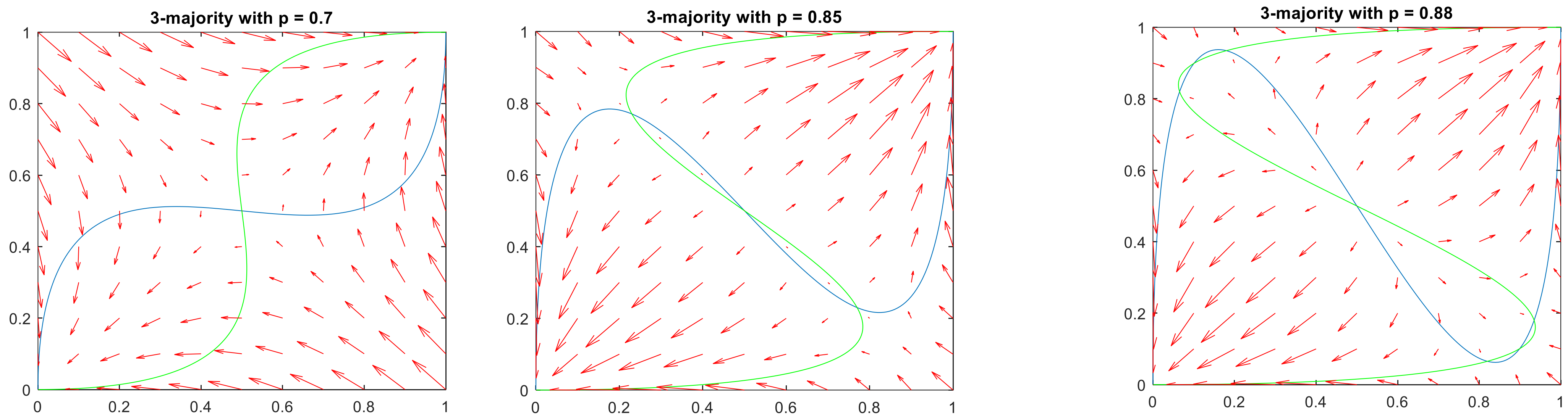
0 δ' δ^* δ'' 1

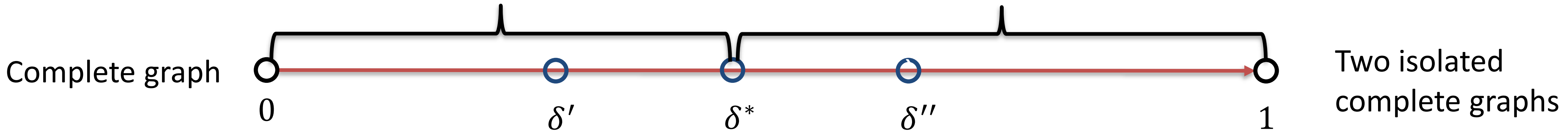
Two isolated complete graphs

Our Dichotomy Theorem

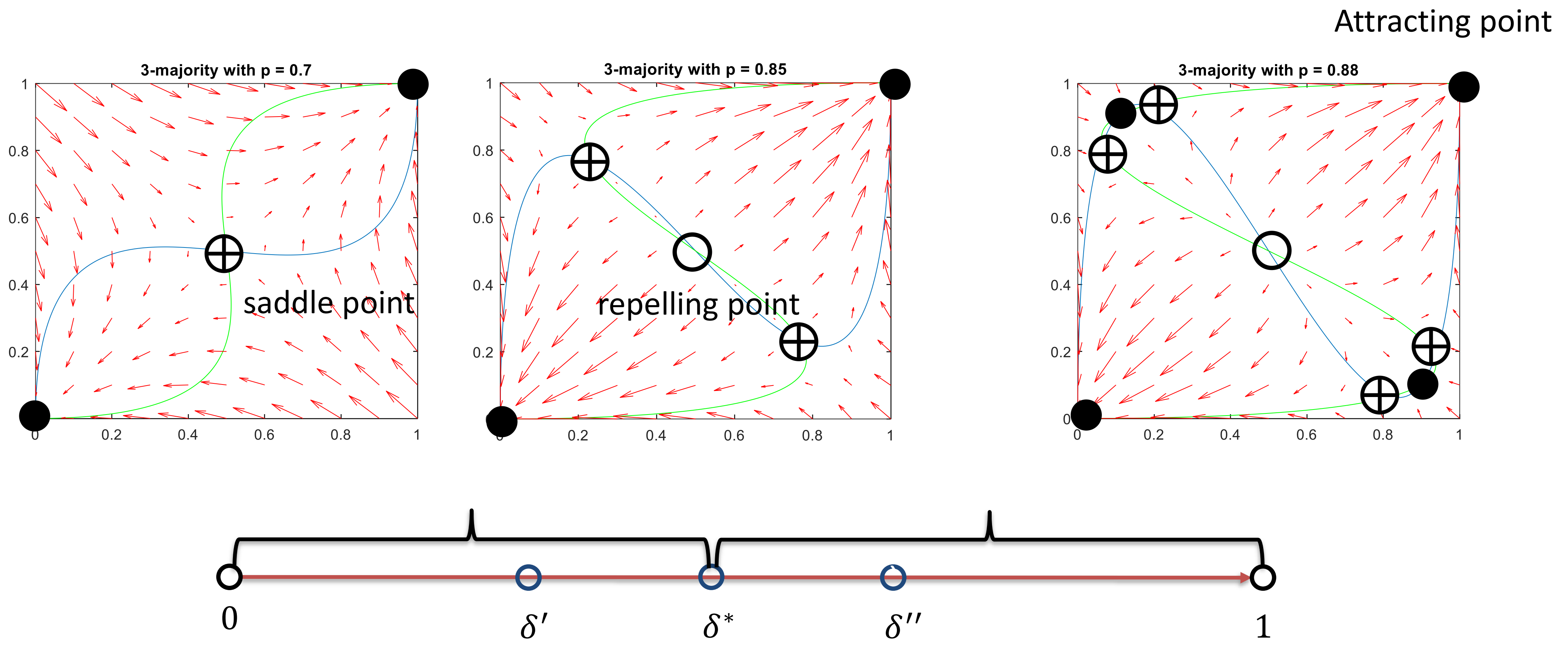


Our Dichotomy Theorem



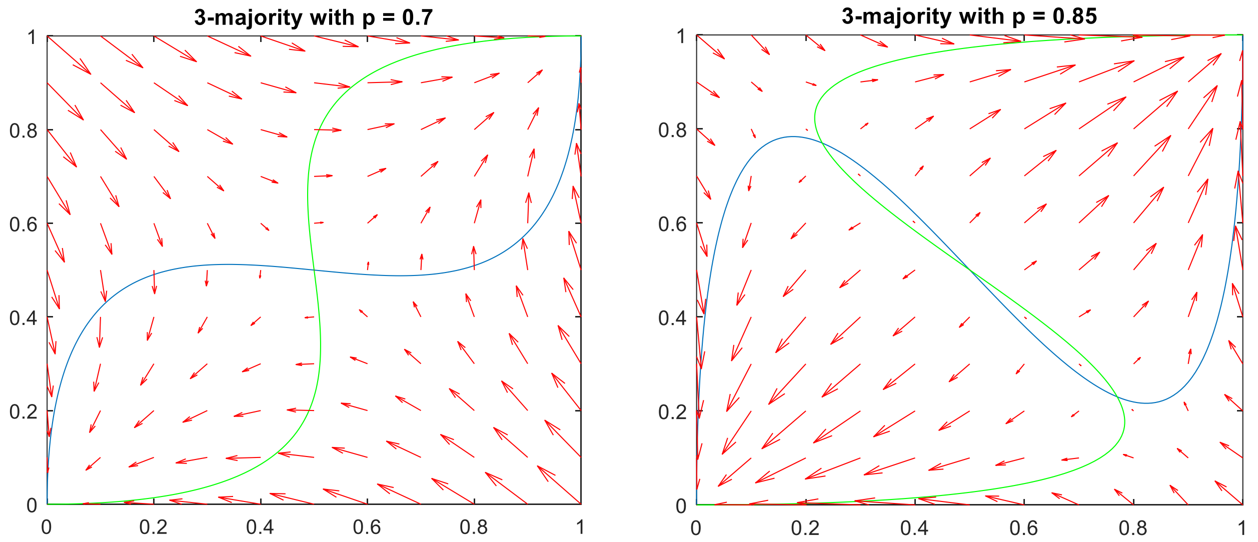


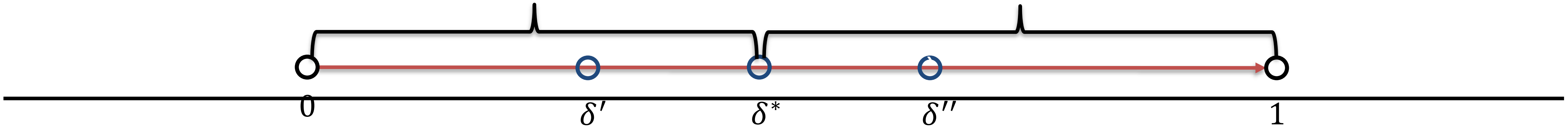
Our Dichotomy Theorem



Fast consensus

$\vec{X}_{k+1} - \vec{X}_k = \frac{1}{n}(F_{ND}(\vec{X}_k) + U(\vec{X}_k))$ reach an attracting fixed point in $O(n \log n)$





Question?