Find Your Place: Simple Distributed Algorithms for Community Detection

Emanuele Natale[†]

joint work with

Luca Becchetti[†], Andrea Clementi[*],
Francesco Pasquale[†] and Luca Trevisan[⋆]

[†] SAPIENZA UNIVERSITÀ DI ROMA
[⋆] UNIVERSITA' DEGLI STUDI DI ROMA TorVergata
[*] Berkeley UNIVERSITY OF CALIFORNIA

Workshop on
Random Processes in Discrete Structures
30 August - 2 September, 2016
Warwick, UK

[⋆]preprint at **goo.gl/aqZmCD**

Dynamics

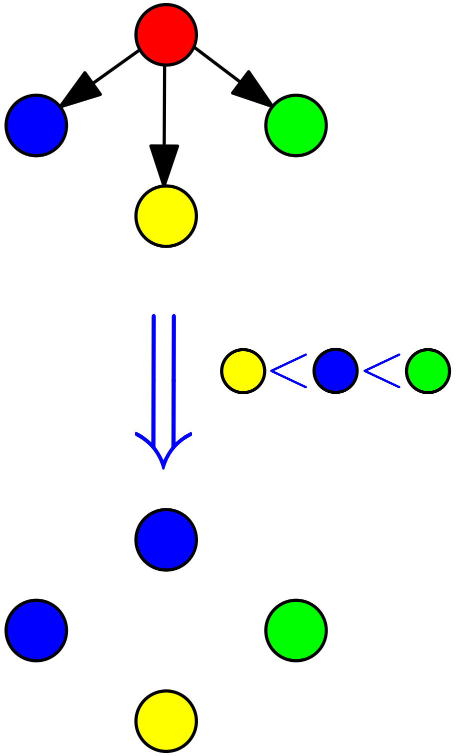
Dynamics: For every graph, agent and round, states are updated according to fixed rule of current state and symmetric function of states of neighbors.

Dynamics

Dynamics: For every graph, agent and round, states are updated according to fixed rule of current state and symmetric function of states of neighbors.

Examples of Dynamics:

- 3-Median dynamics [Doerr et al. '11]

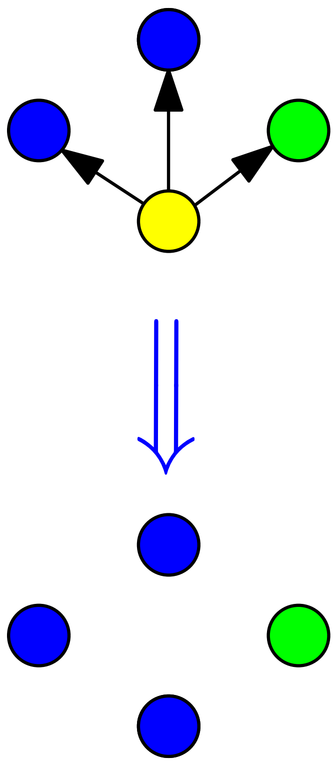


Dynamics

Dynamics: For every graph, agent and round, states are updated according to fixed rule of current state and symmetric function of states of neighbors.

Examples of Dynamics:

- 3-Median dynamics
  [Doerr et al. '11]
- 3-Majority dynamics
  [Becchetti et al. '14, '16]

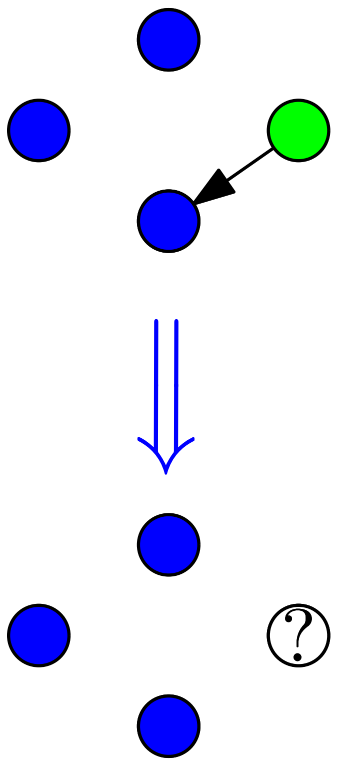


Dynamics

Dynamics: For every graph, agent and round, states are updated according to fixed rule of current state and symmetric function of states of neighbors.

Examples of Dynamics:

- 3-Median dynamics
  [Doerr et al. '11]
- 3-Majority dynamics
  [Becchetti et al. '14, '16]
- Undecided-state dynamics
  [Becchetti et al. '15]

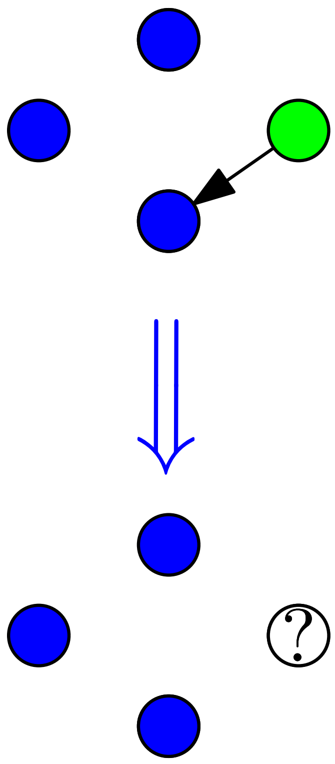


Dynamics

Dynamics: For every graph, agent and round, states are updated according to fixed rule of current state and symmetric function of states of neighbors.

Examples of Dynamics:

- 3-Median dynamics
  [Doerr et al. '11]
- 3-Majority dynamics
  [Becchetti et al. '14, '16]
- Undecided-state dynamics
  [Becchetti et al. '15]

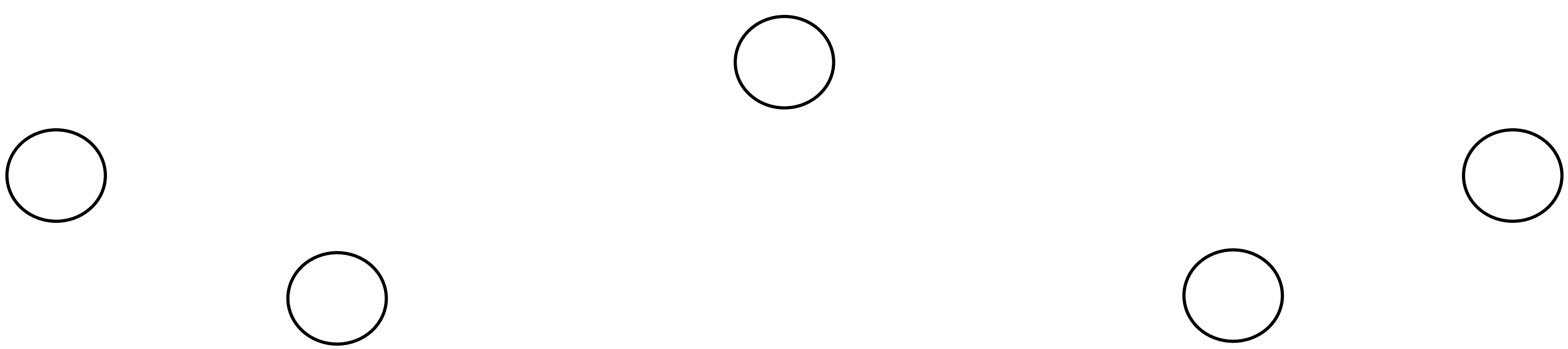


Can dynamics solve a problem non-trivial in centralized setting?

The Average Dynamics

Al nodes at the same time:

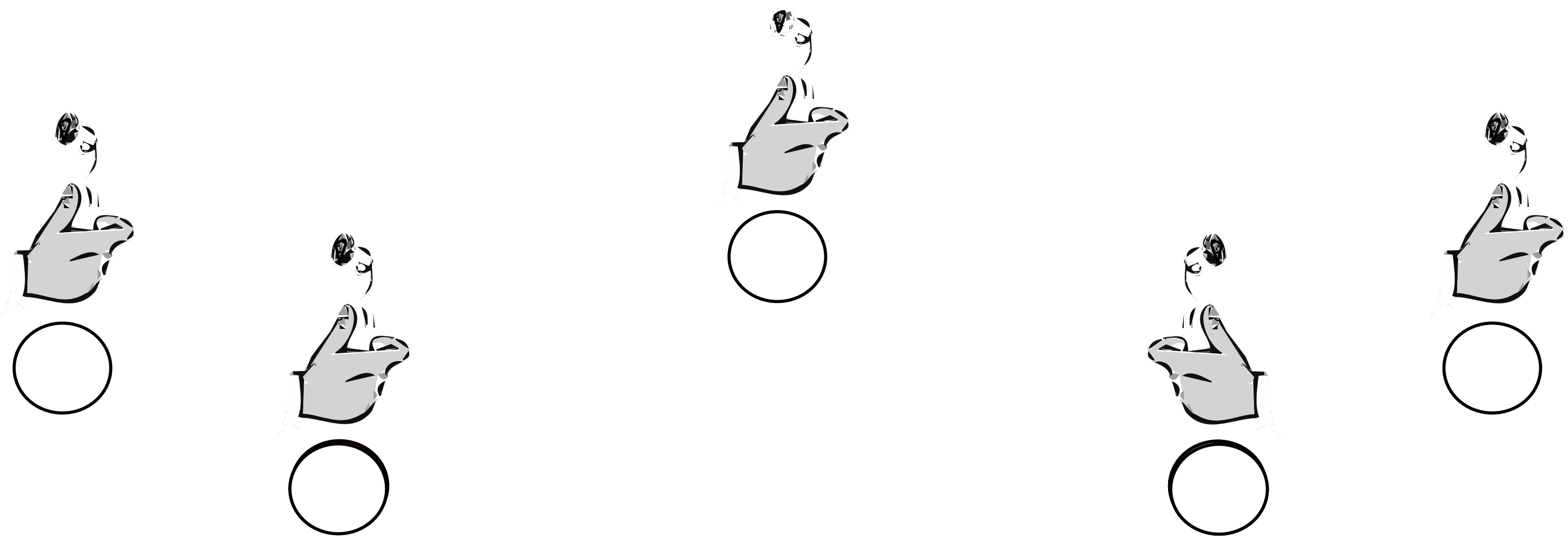
- At $t = 0$, randomly pick value $x^{(t)} \in \{\mathbf{+1}, \mathbf{-1}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.



The Average Dynamics

Al nodes at the same time:

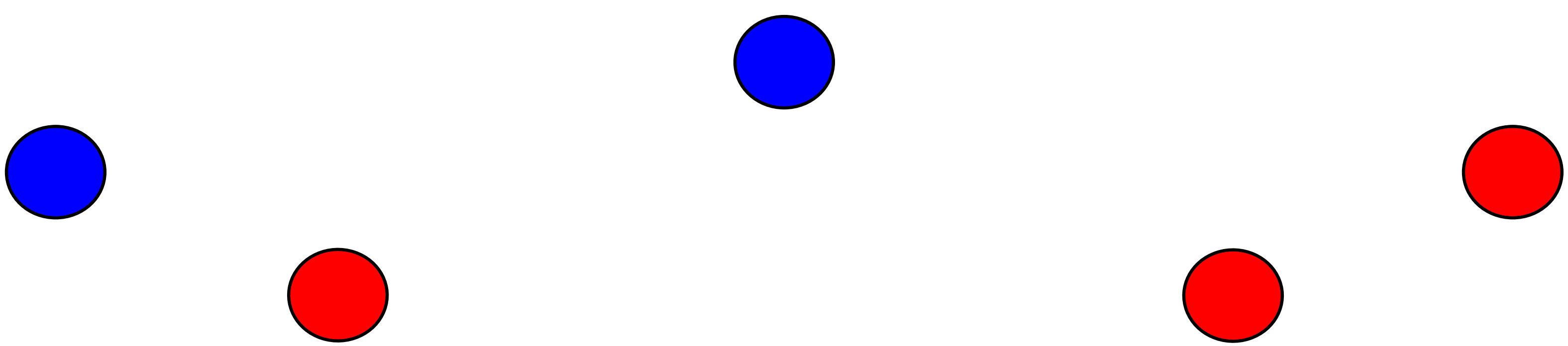
- At $t = 0$, randomly pick value $x^{(t)} \in \{\mathbf{+1}, \mathbf{-1}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.



The Average Dynamics

Al nodes at the same time:

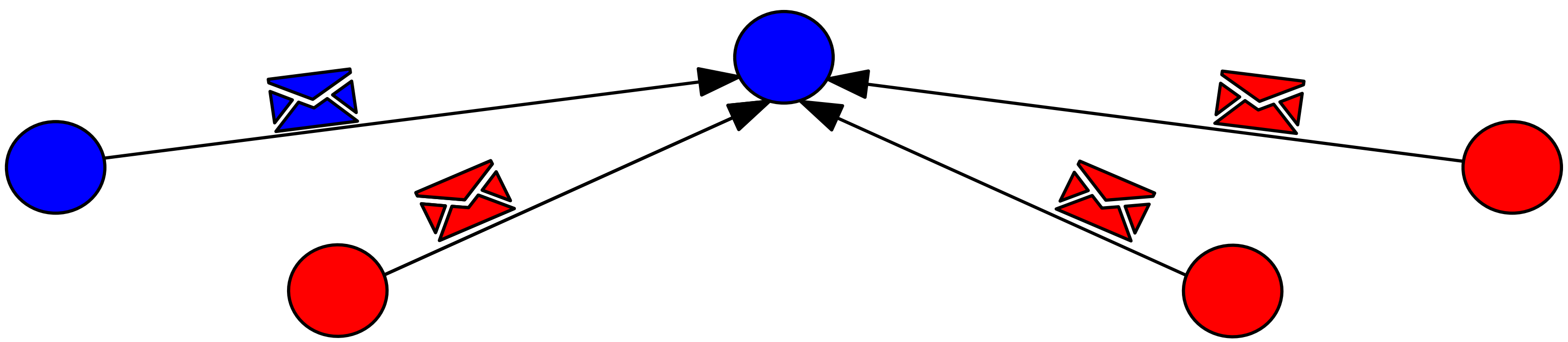
- At $t = 0$, randomly pick value $x^{(t)} \in \{\mathbf{+1}, \mathbf{-1}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.



The Average Dynamics

Al nodes at the same time:

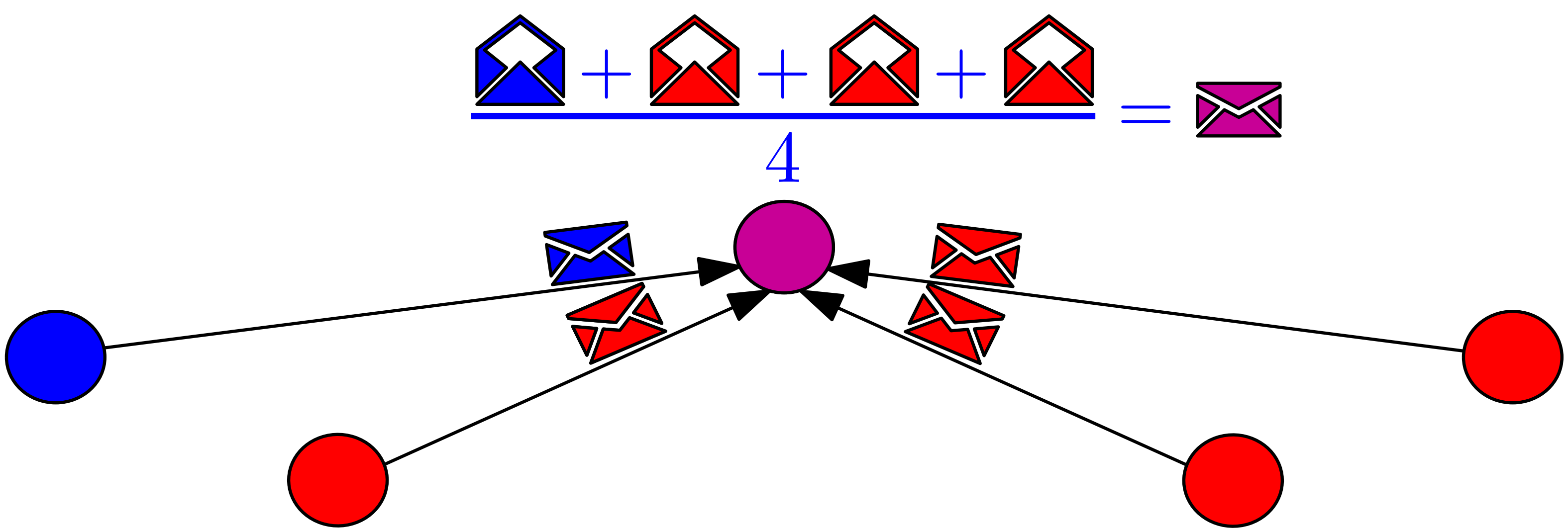
- At $t = 0$, randomly pick value $x^{(t)} \in \{\mathbf{+1}, \mathbf{-1}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.



The Average Dynamics

Al nodes at the same time:

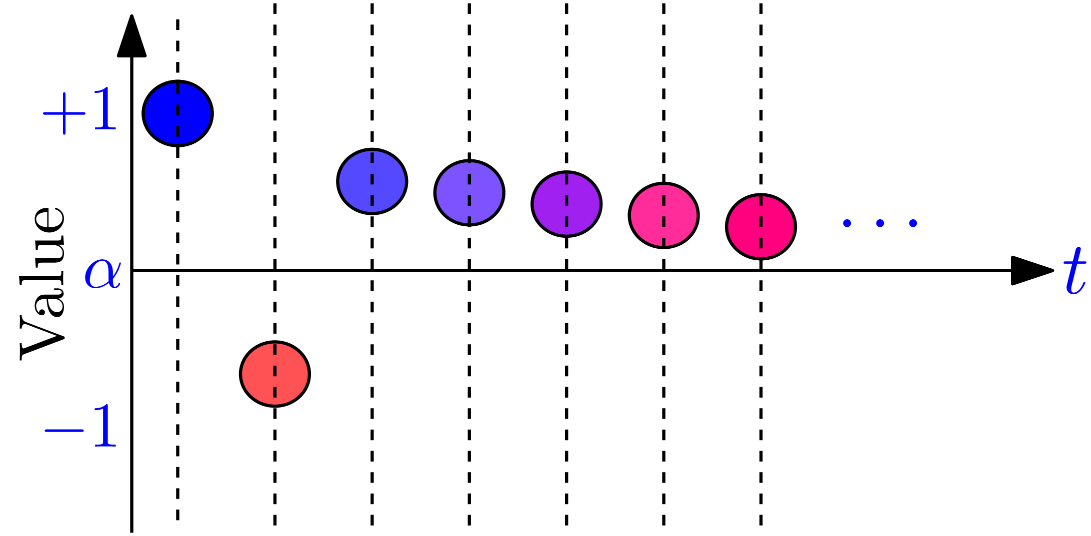
- At $t = 0$, randomly pick value $x^{(t)} \in \{\mathbf{+1}, \mathbf{-1}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.

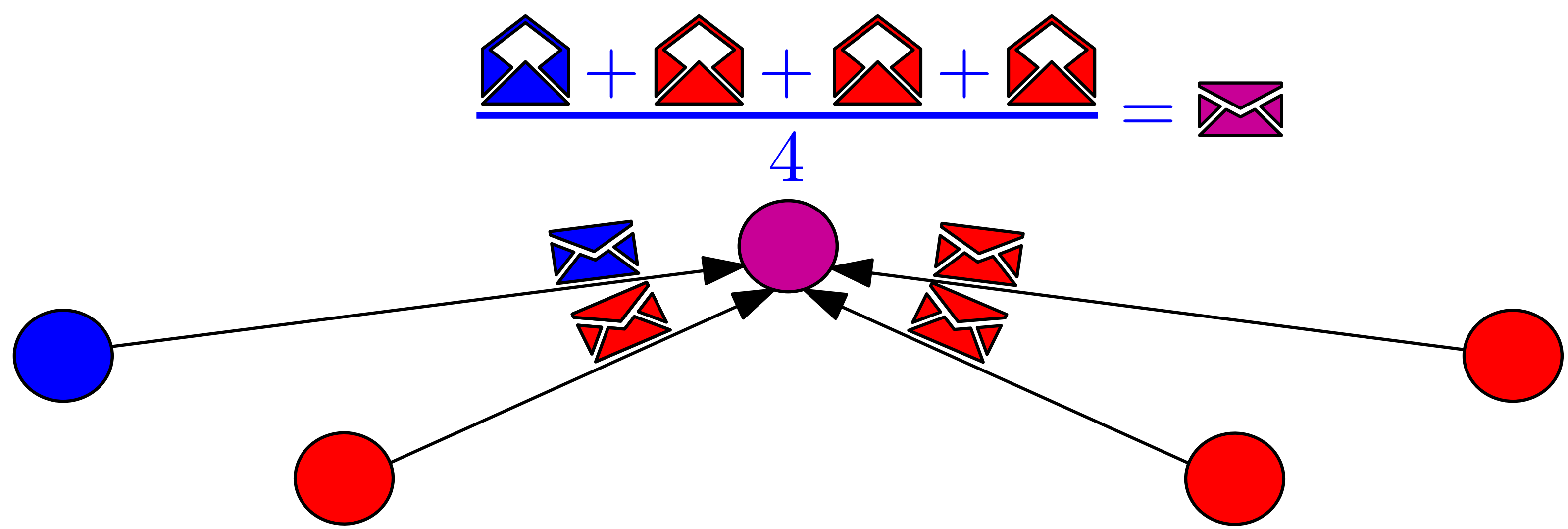


The Average Dynamics

Al nodes at the same time:

- At $t = 0$, randomly pick value $x^{(t)} \in \{\mathbf{+1}, \mathbf{-1}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.

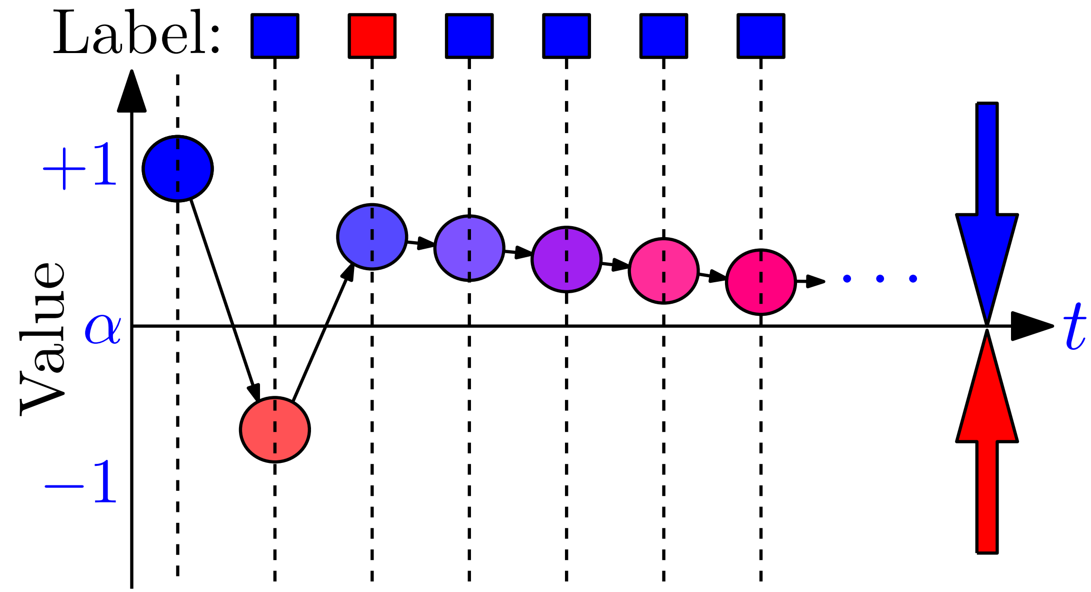


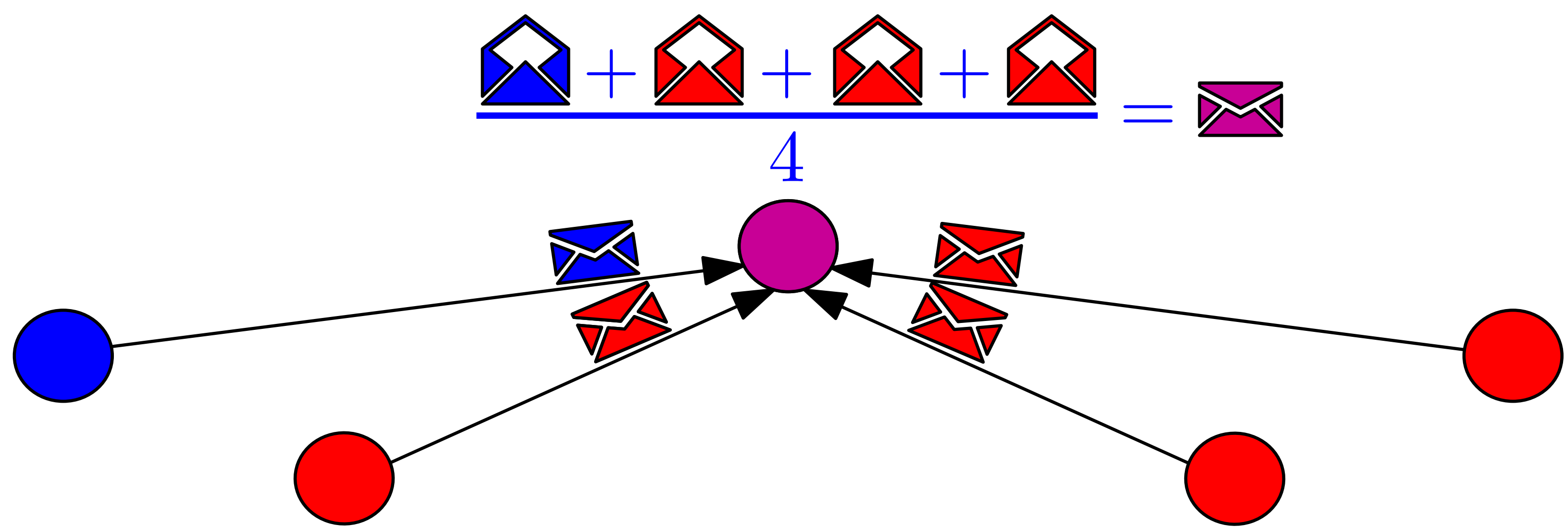


The Average Dynamics

Al nodes at the same time:

- At $t = 0$, randomly pick value $x^{(t)} \in \{\mathbf{+1}, \mathbf{-1}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.





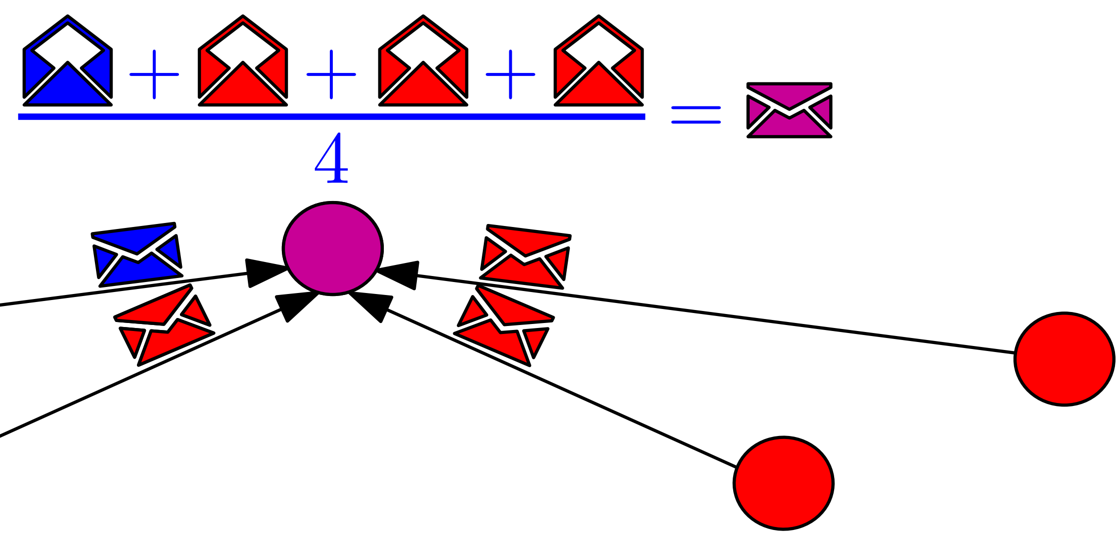
The Average Dynamics

Al nodes at the same time:

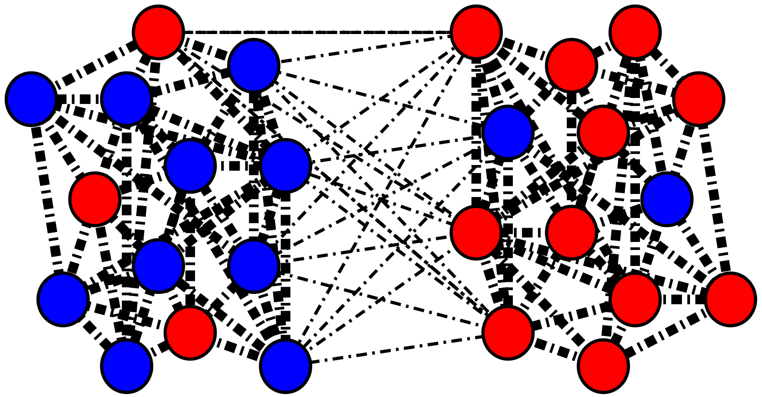
- At $t = 0$, randomly pick value $x^{(t)} \in \{\mathbf{+1}, \mathbf{-1}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.

Well studied process [Shah '09]:

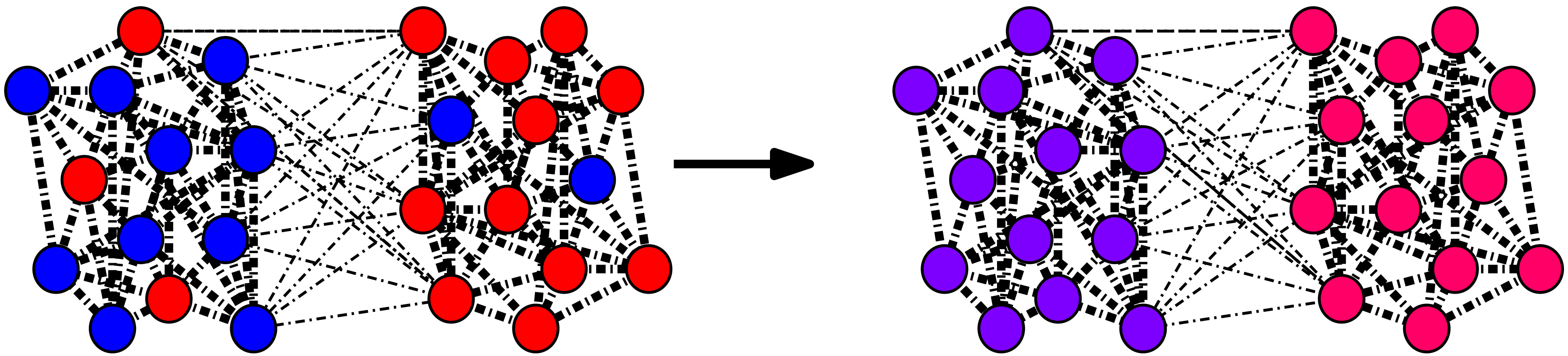
- Converges to (weighted) global average of initial values,
- Convergence time $=$ mixing time of G,
- Important applications in fault-tolerant self-stabilizing consensus.



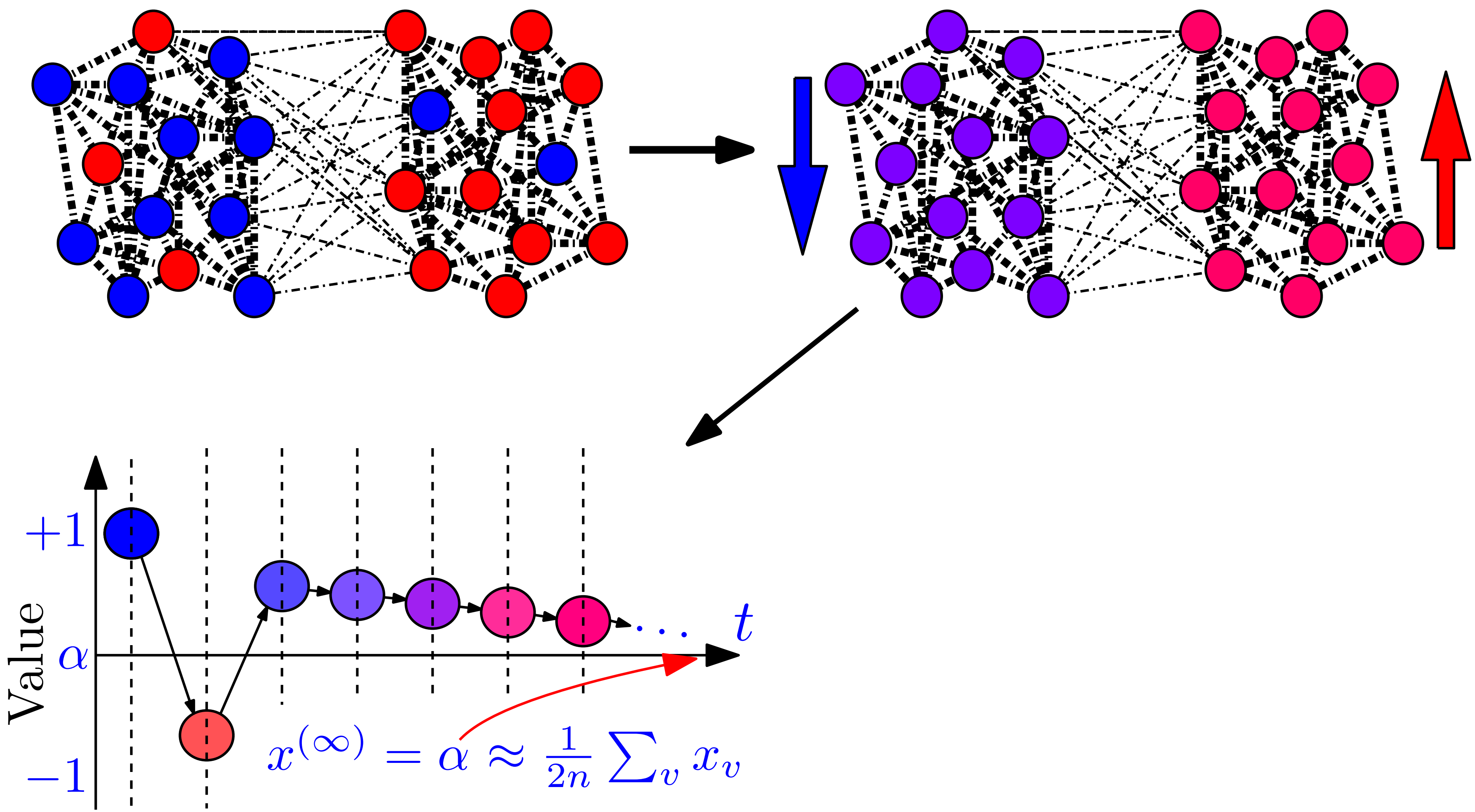
Our Results



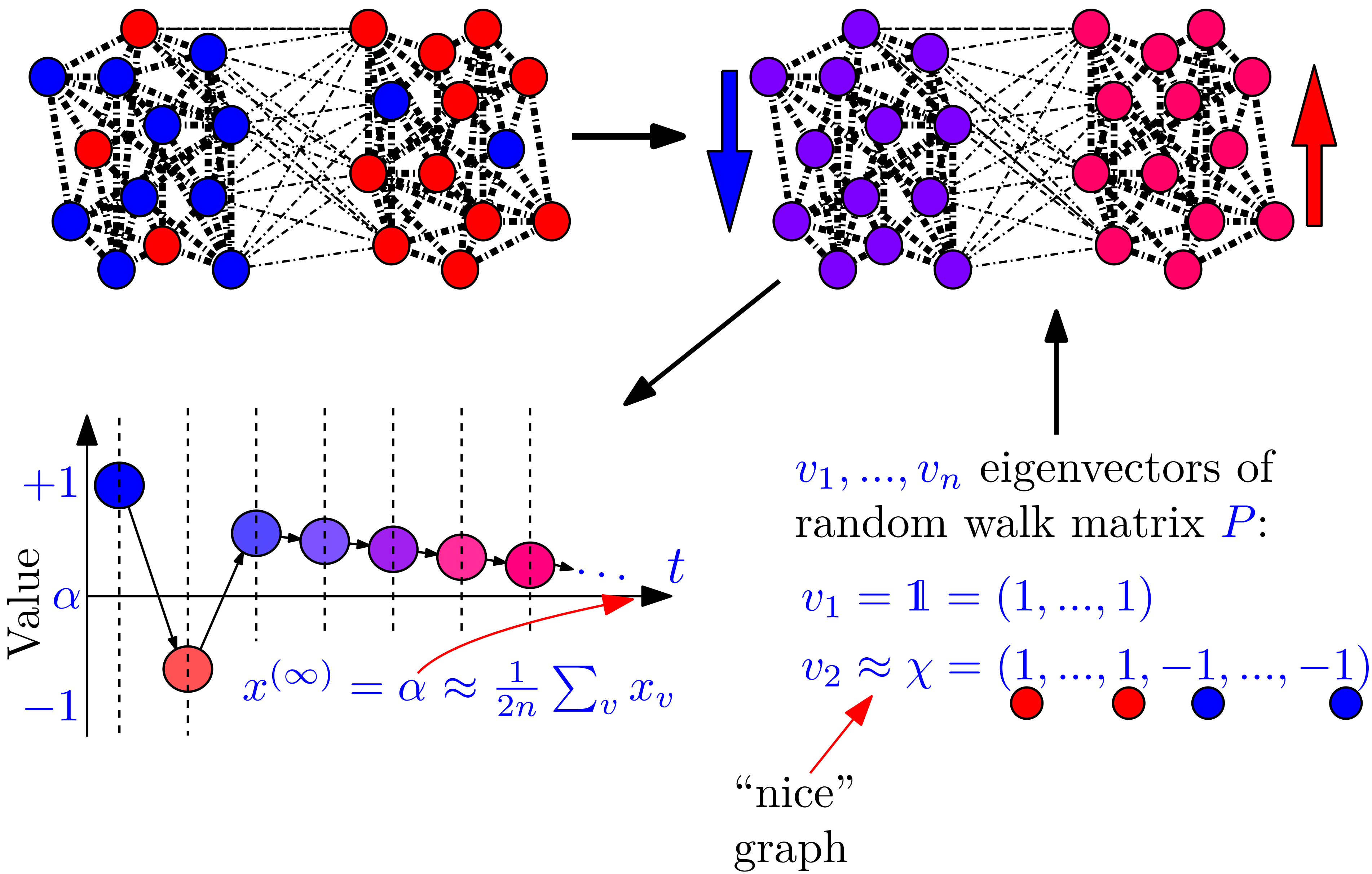
Our Results



Our Results

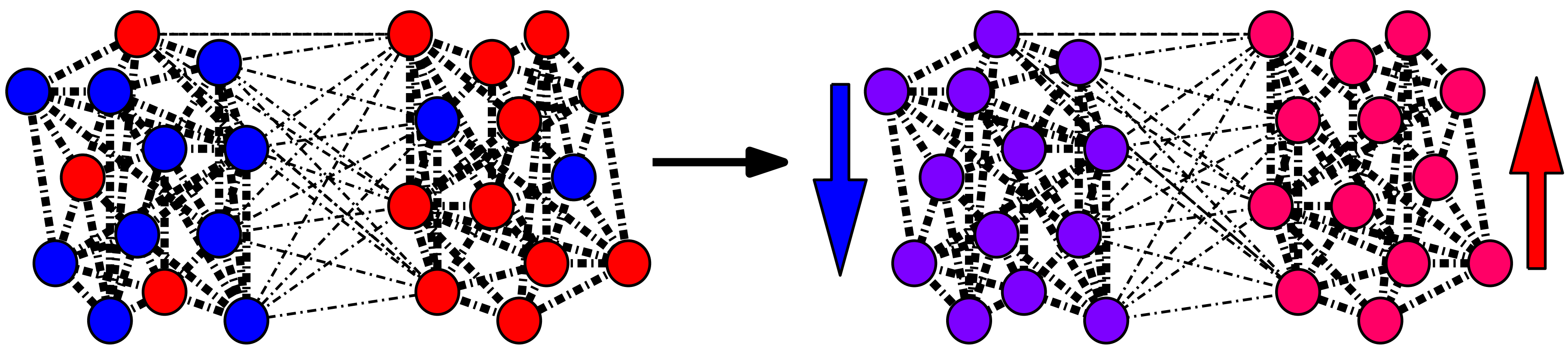


Our Results



Value
+1
α
−1
t
$x^{(\infty)} = \alpha \approx \frac{1}{2n}\sum_v x_v$
$v_1, ..., v_n$ eigenvectors of
random walk matrix P:
$v_1 = \mathbb{1} = (1, ..., 1)$
$v_2 \approx \chi = (1, ..., 1, -1, ..., -1)$
“nice”
graph

Our Results

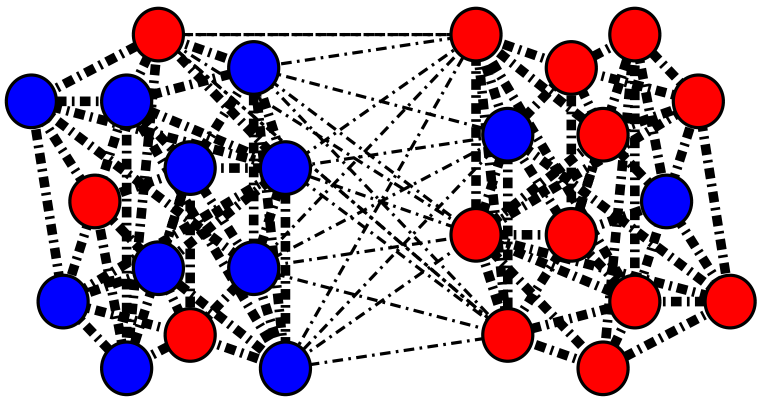


(Informal) Theorem. $G = (V_1 \dot{\bigcup} V_2, E)$ s.t.
i) $\chi = \mathbf{1}_{V_1} - \mathbf{1}_{V_2}$ close to right-eigenvector of eigenvalue λ_2 of transition matrix of G, and
ii) gap between λ_2 and $\lambda = \max\{\lambda_3, |\lambda_n|\}$ sufficiently large, then
Averaging (approximately) identifies (V_1, V_2).

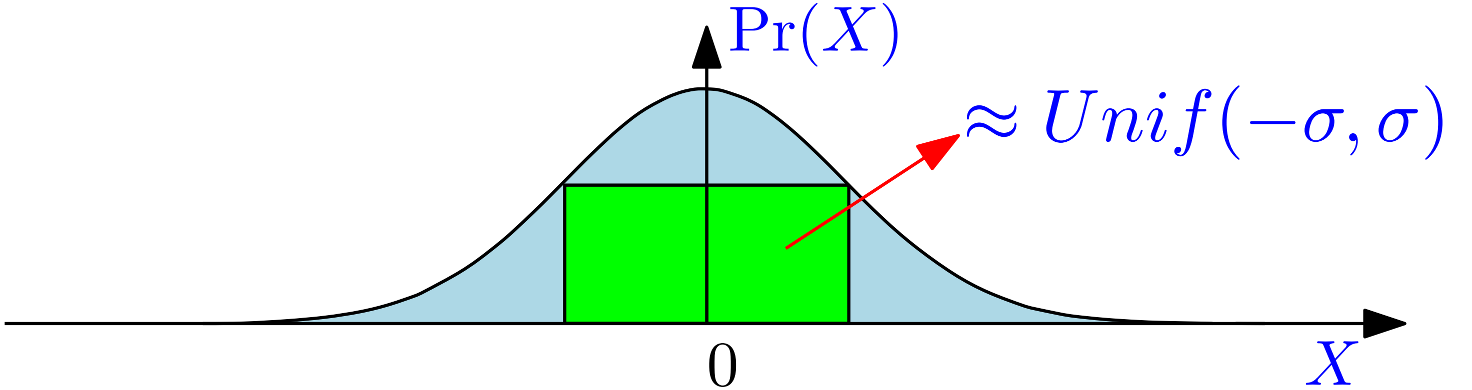
Properties of the Averaging Dynamics

Al nodes at the same time:

- At $t = 0$, randomly pick value $x^{(t)} \in \{\textbf{blue}, \textbf{red}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.



$$\Pr\left(\left|\sum_{v\in V_1}\mathbf{x}(v) - \sum_{v\in V_2}\mathbf{x}(v)\right| > n^{\epsilon}\right) \geq 1 - n^{\Omega(1)} \text{ (w.h.p.)}$$



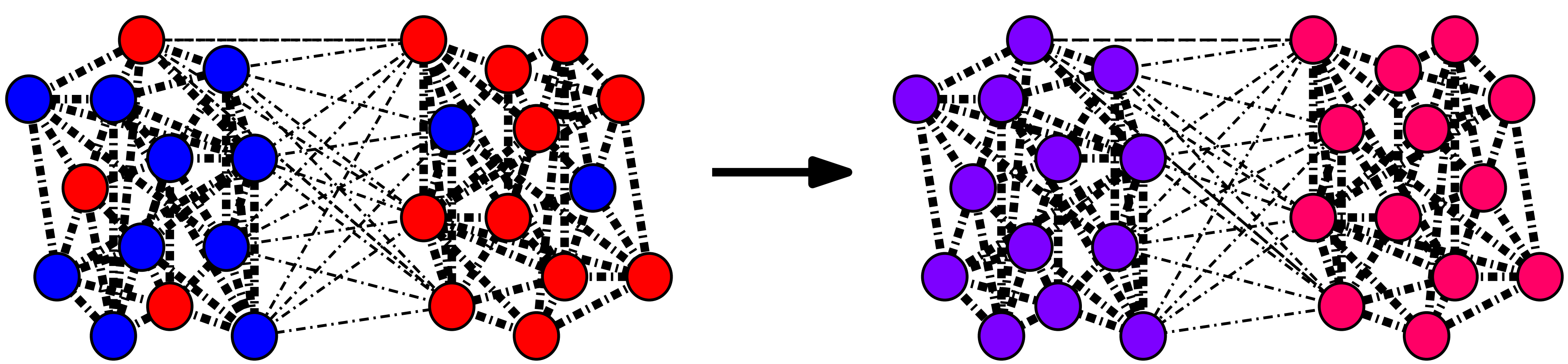
Properties of the Averaging Dynamics

Al nodes at the same time:

- At $t = 0$, randomly pick value $x^{(t)} \in \{\textbf{blue}, \textbf{red}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.

Closely related to simple random walk on G:
$Y_v^{(t)} :=$ position at time t of simple random walk starting from v

$$\implies x^{(t)}(v) = \mathbb{E}[x^{(0)}(Y_v^{(t)})]$$



Properties of the Averaging Dynamics

Al nodes at the same time:

- At $t = 0$, randomly pick value $x^{(t)} \in \{\textbf{blue}, \textbf{red}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.

Features:

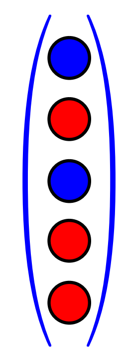
- No explicit eigenvector computation
- Implicit "simulation" of power method

$A = (\mathbb{1}_{((u,v)\in E)})_{u,v\in V}$ adjacency matrix of G

D diagonal matrix of node degrees in G

$P = D^{-1}A$ transition matrix of random walk

Averaging is a **linear** dynamics

$\mathbf{x}^{(t)} =$ 

$$\mathbf{x}^{(t)} = P \cdot \mathbf{x}^{(t-1)} = P^t \cdot \mathbf{x}^{(0)}$$

Properties of the Averaging Dynamics

Al nodes at the same time:

- At $t = 0$, randomly pick value $x^{(t)} \in \{\mathbf{blue}, \mathbf{red}\}$.
- Then, at each round
  1. Set color $x^{(t)}$ to average of neighbors,
  2. Set label to **blue** if $x^{(t)} < x^{(t-1)}$, **red** otherwise.

Remove projection on first eigenspace $\Longrightarrow$ running time depending on λ_2/λ

Bottleneck of mixing time for spectral methods:

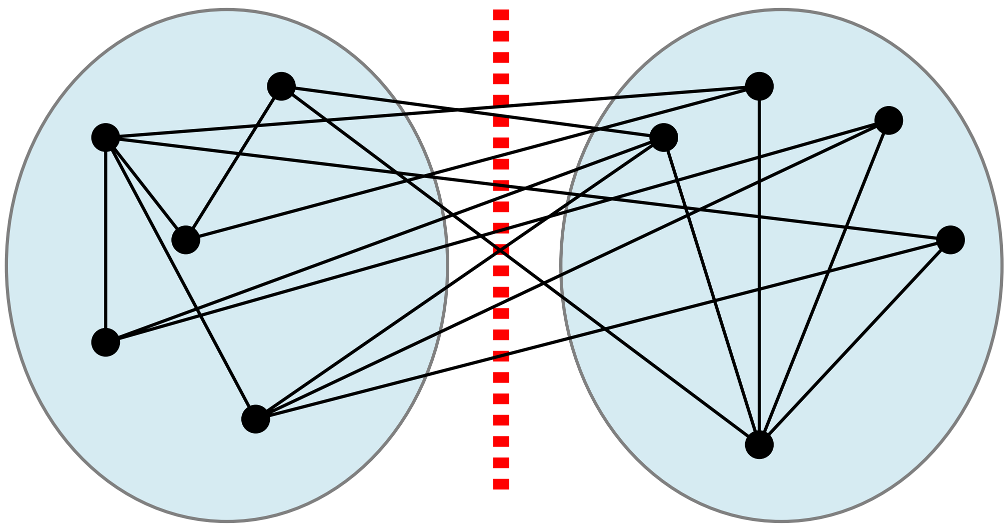
Distributed computation of second eigenvector [Kempe & McSherry '08]: $\mathcal{O}(\tau_{mix} \log^2 n)$.

Community Detection as Minimum Bisection

Minimum Bisection Problem.

Input: a graph G with $2n$ nodes.

Output: $S = \arg\min_{\substack{S \subset V \\ |S| = n}} E(S, V - S)$.

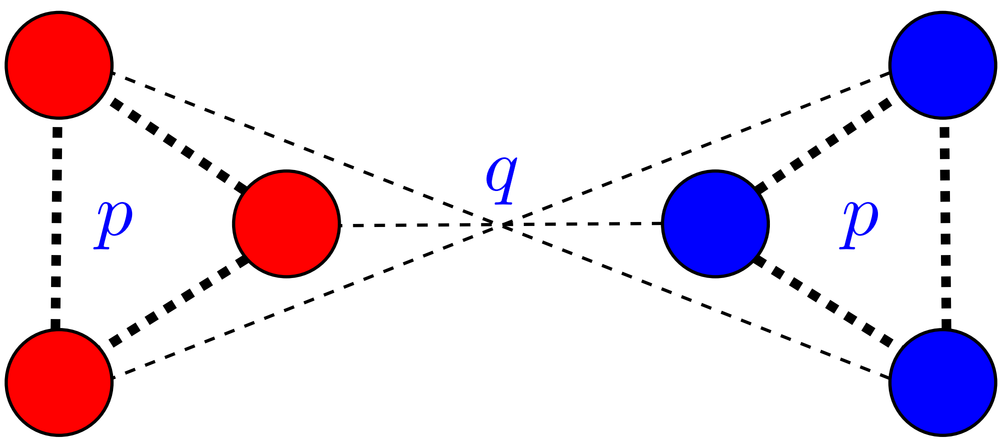


[Garey, Johnson, Stockmeyer '76]:

Min-Bisection is *NP-Complete*.

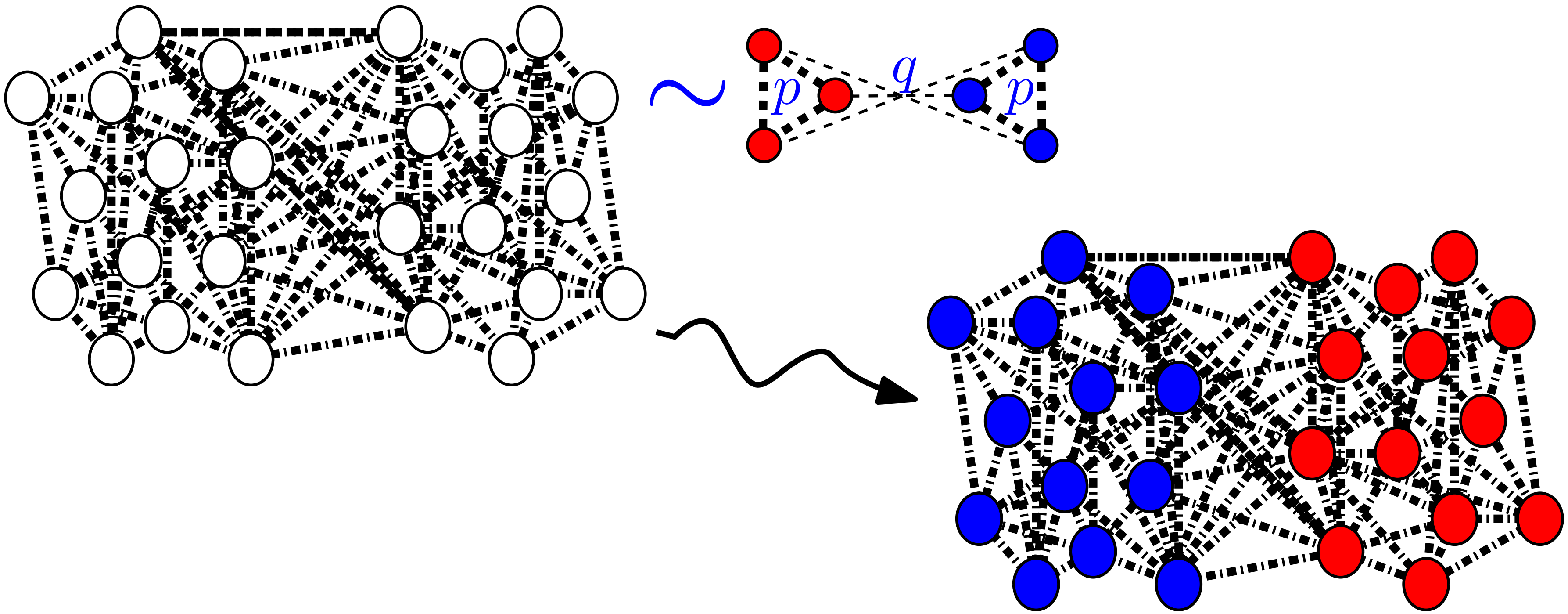
The Stochastic Block Model

Stochastic Block Model (SBM). Two "communities" of equal size V_1 and V_2, each edge inside a community included with probability $p = \frac{a}{n}$, each edge across communities included with probability $q = \frac{b}{n} < p$.



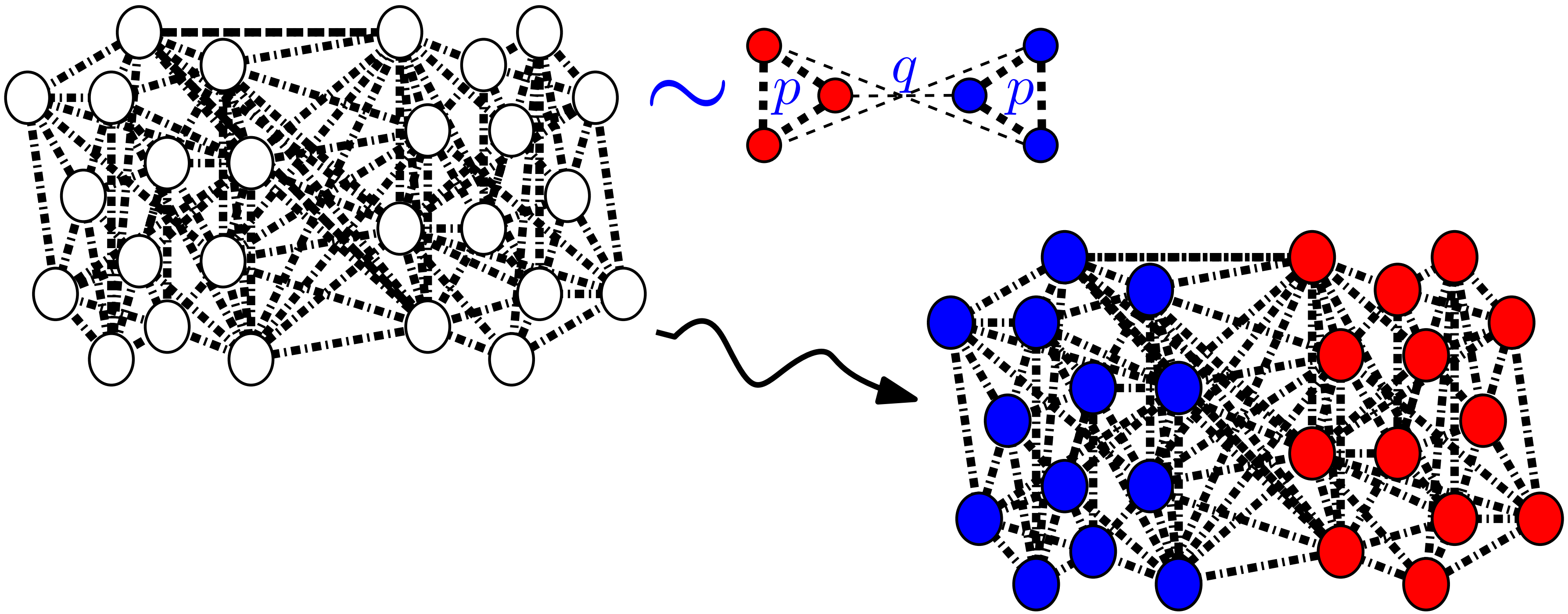
The Stochastic Block Model

Reconstruction problem. Given graph generated by SBM, find original partition.



The Stochastic Block Model

Reconstruction problem. Given graph generated by SBM, find original partition.

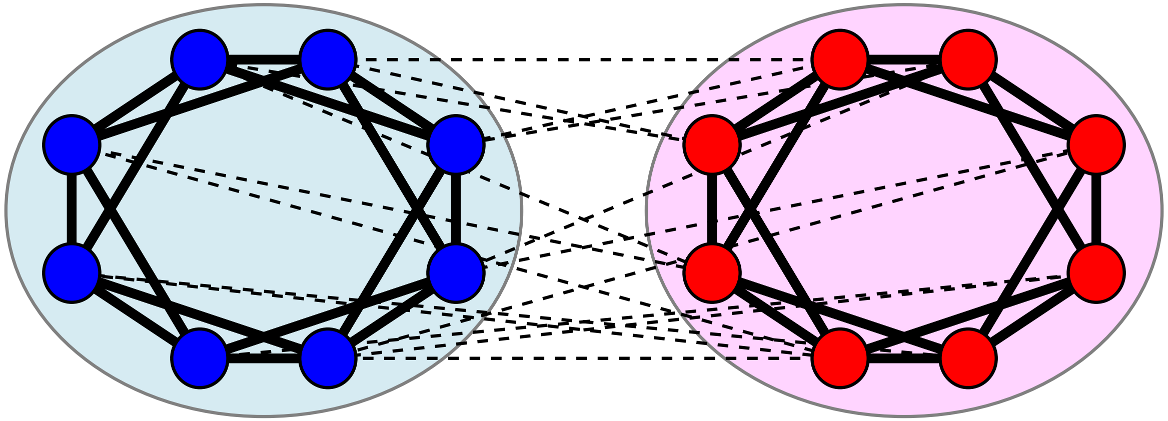


$\lambda_2(P) \approx \frac{a-b}{d} \implies$ mixing time of a random walk on $\mathcal{G}_{2n,\frac{a}{n},\frac{b}{n}}$ is $\geq \frac{1}{1-\lambda_2} \approx \frac{a+b}{2b}$.

Regular Stochastic Block Model

Regular SBM (RSBM) [Brito et al. SODA'16]. A graph $G = (V_1 \dot{\bigcup} V_2, E)$ s.t.

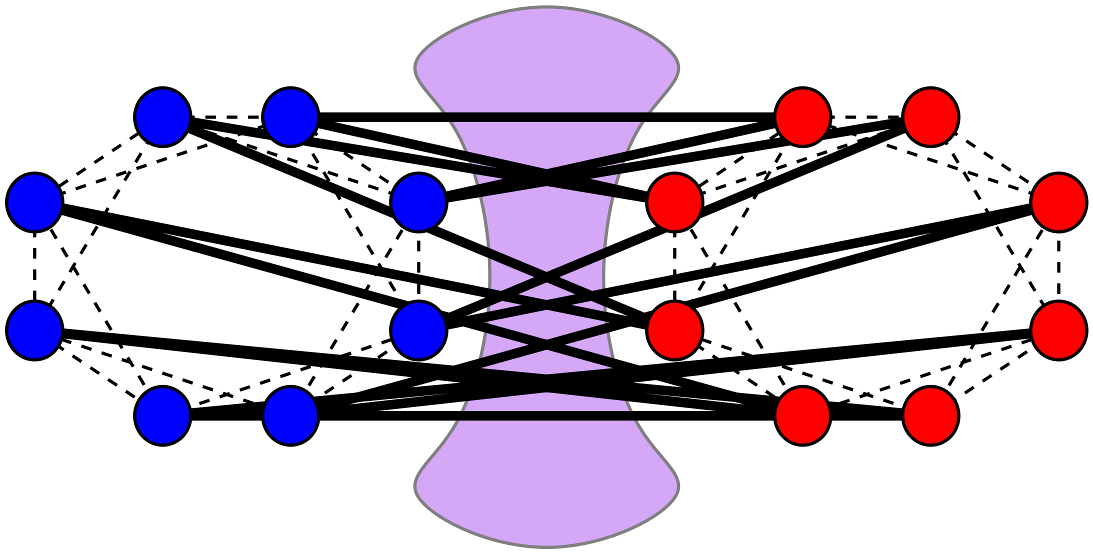
- $|V_1| = |V_2|$,
- $G\big|_{V_1}, G\big|_{V_2} \sim$ random a-regular graphs
- $G\big|_{E(V_1,V_2)} \sim$ random b-regular bipartite graph.



Regular Stochastic Block Model

Regular SBM (RSBM) [Brito et al. SODA'16]. A graph $G = (V_1 \dot{\bigcup} V_2, E)$ s.t.

- $|V_1| = |V_2|$,
- $G\big|_{V_1}, G\big|_{V_2} \sim$ random a-regular graphs
- $G\big|_{E(V_1,V_2)} \sim$ random b-regular bipartite graph.

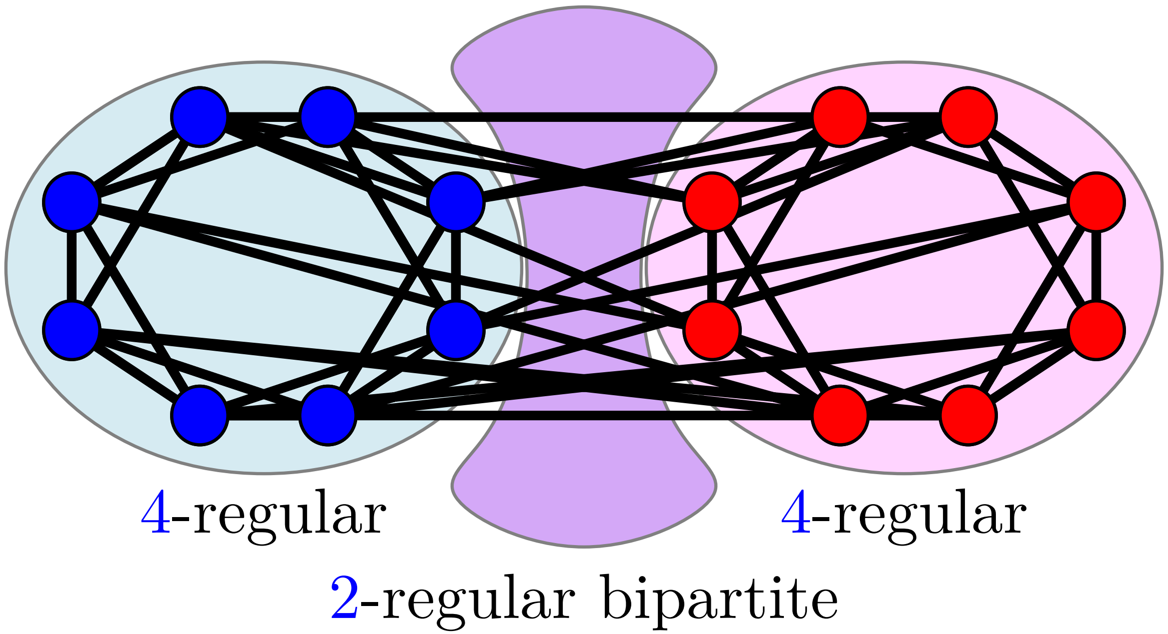


2-regular bipartite

Regular Stochastic Block Model

Regular SBM (RSBM) [Brito et al. SODA'16]. A graph $G = (V_1 \dot{\bigcup} V_2, E)$ s.t.

- $|V_1| = |V_2|$,
- $G\big|_{V_1}, G\big|_{V_2} \sim$ random a-regular graphs
- $G\big|_{E(V_1,V_2)} \sim$ random b-regular bipartite graph.



When is Reconstruction Possible?

[Decelle, Massoulie, Mossel, Brito, Abbe et al.]:
Reconstruction is possible iff

- $a - b > 2\sqrt{d}$ in SBM (weak)
- $a - b > 2(\sqrt{a} - \sqrt{b})\sqrt{b} + 2\log n$ in SBM (strong)
- $a - b > 2\sqrt{d-1}$ in RSBM (strong)

Linearizations of *Belief Propagation*, advanced spectral methods (power and Lanczos method), SDP.

When is Reconstruction Possible?

[Decelle, Massoulie, Mossel, Brito, Abbe et al.]:
Reconstruction is possible iff

- $a - b > 2\sqrt{d}$ in SBM (weak)
- $a - b > 2(\sqrt{a} - \sqrt{b})\sqrt{b} + 2\log n$ in SBM (strong)
- $a - b > 2\sqrt{d-1}$ in RSBM (strong)

Linearizations of *Belief Propagation*, advanced spectral methods (power and Lanczos method), SDP.

Not a dynamics: nonlinear, different messages to different neighbors

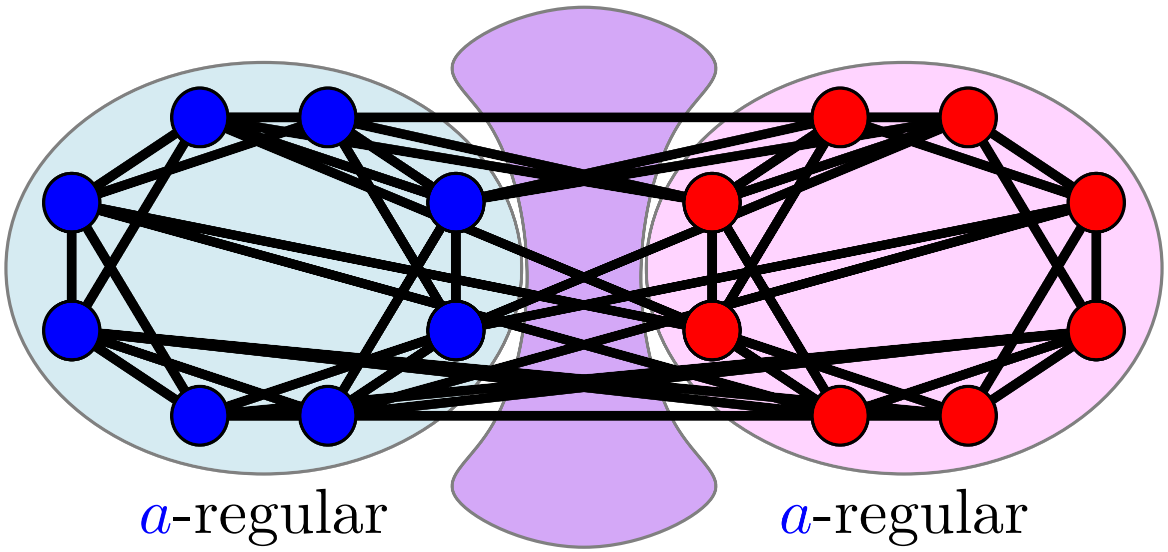
Centralized, not easy to make distribute

Regular Clustered and Clustered Graphs

$(2n, d, b)$-clustered Regular Graph.

A graph $G = (V_1 \dot{\bigcup} V_2, E)$ s.t.

- $|V_1| = |V_2|$,
- G is d regular,
- each $v \in V_i$ has b neighbors in V_{3-i}.



No randomness!

Regular Clustered and Clustered Graphs

$(2n, d, b)$**-clustered Regular Graph.**
A graph $G = (V_1 \dot{\bigcup} V_2, E)$ s.t.

- $|V_1| = |V_2|$,
- G is d regular,
- each $v \in V_i$ has b neighbors in V_{3-i}.

Thm. If $G\big|_{V_1}$, $G\big|_{V_2}$ expanders and $\lambda_2/\lambda > 1$ (e.g. if $b \ll d/2$), averaging produces strong reconstruction in $\mathcal{O}(\log n)$ rounds.

Regular Clustered and Clustered Graphs

$(2n, d, b)$**-clustered Regular Graph.**
A graph $G = (V_1 \dot{\bigcup} V_2, E)$ s.t.

- $|V_1| = |V_2|$,
- G is d regular,
- each $v \in V_i$ has b neighbors in V_{3-i}.

Thm. If $G\big|_{V_1}$, $G\big|_{V_2}$ expanders and $\lambda_2/\lambda > 1$ (e.g. if $b \ll d/2$), averaging produces strong reconstruction in $\mathcal{O}(\log n)$ rounds.

RSBM is $(2n, d, b)$-clustered regular
with $G\big|_{V_1}$, $G\big|_{V_2}$ expanders w.h.p. $\implies$
Cor. Strong reconstruction ($a - b > 2\sqrt{d-1}$)

Regular Clustered and Clustered Graphs

$(2n, d, b, \gamma)$**-clustered Graph.**
A graph $G = (V_1 \dot{\bigcup} V_2, E)$ s.t.

- $|V_1| = |V_2|$,
- every node has degree $d \pm \gamma d$
- each $v \in V_i$ has $b \pm \gamma d$ neighbors in V_{3-i}.

Thm. If $\min\{\lambda_2, \frac{a-b}{d}\} > \lambda$ and $\gamma = \mathcal{O}(\frac{a-b}{d} - \lambda_3)$
$\implies \mathcal{O}(\gamma^2/(\frac{a-b}{d} - \lambda_3)^2)$-weak reconstruction.

Regular Clustered and Clustered Graphs

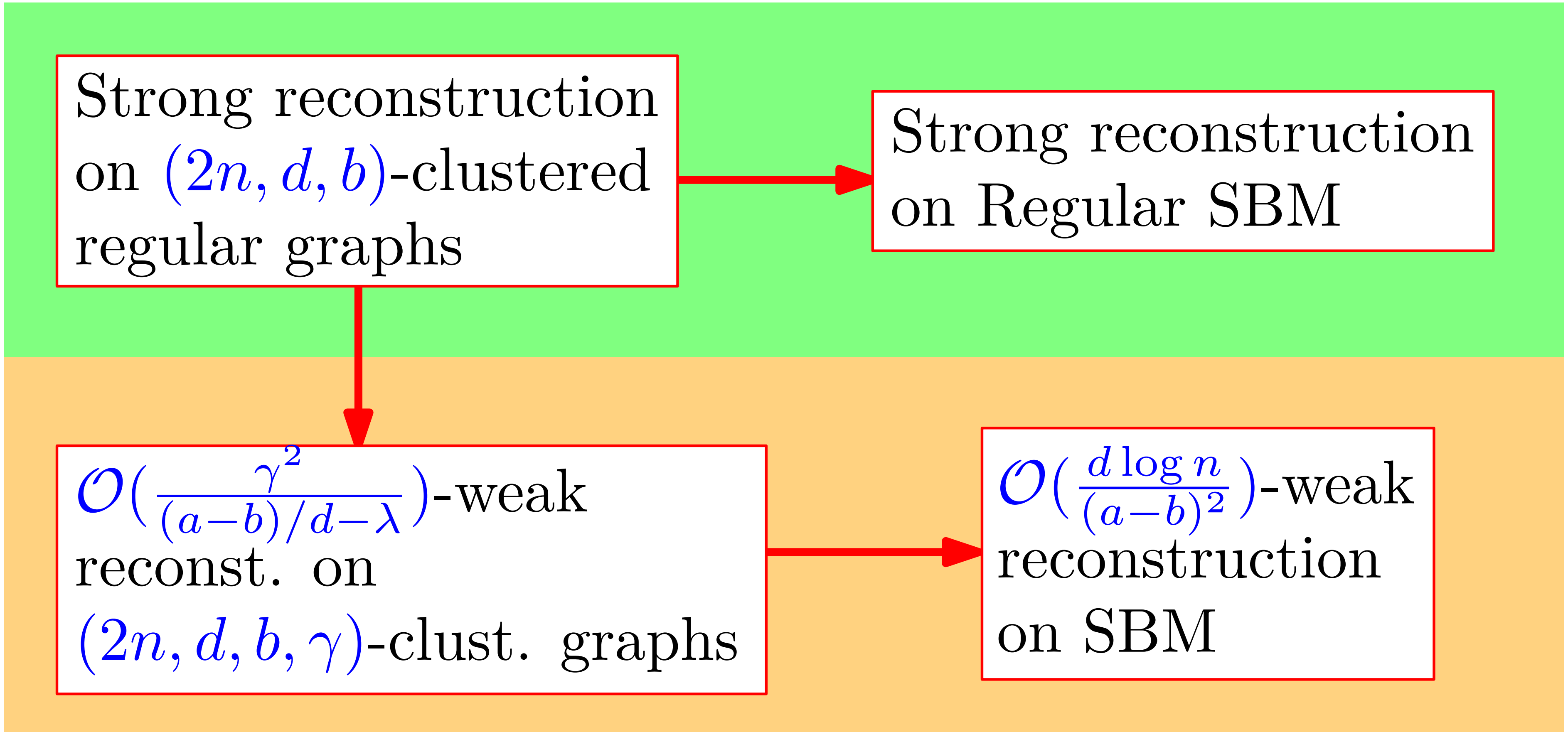
$(2n, d, b, \gamma)$-clustered Graph.
A graph $G = (V_1 \dot{\bigcup} V_2, E)$ s.t.

- $|V_1| = |V_2|$,
- every node has degree $d \pm \gamma d$
- each $v \in V_i$ has $b \pm \gamma d$ neighbors in V_{3-i}.

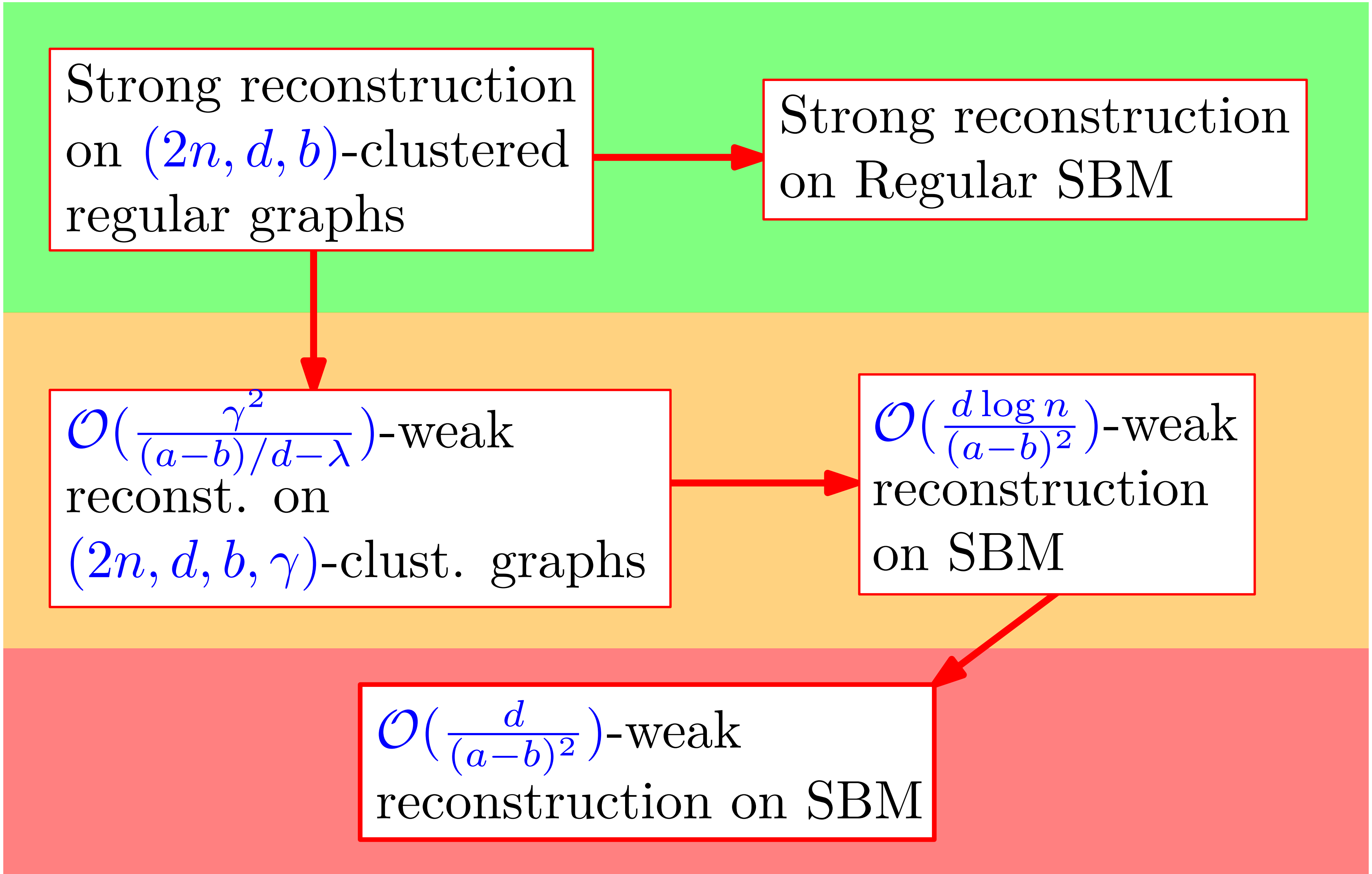
Thm. If $\min\{\lambda_2, \frac{a-b}{d}\} > \lambda$ and $\gamma = \mathcal{O}(\frac{a-b}{d} - \lambda_3)$
$\implies \mathcal{O}(\gamma^2/(\frac{a-b}{d} - \lambda_3)^2)$-weak reconstruction.

Cor. If $a - b > \sqrt{d \log n}$ and $b > \frac{\log n}{n^2}$, SBM is
$(2n, d, b, 6\sqrt{\frac{\log n}{d}})$-clust. with $\min\{\lambda_2, 24\sqrt{\frac{\log n}{d}}\} > \lambda$
w.h.p. $\implies \mathcal{O}\left(\frac{d \log n}{(a-b)^2}\right)$-weak reconstruction.

Analysis: Roadmap



Analysis: Roadmap



Analysis on Regular Graphs

$P = D^{-1}A = \frac{1}{d}A$ → symmetric $\Longrightarrow$ orthonormal eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_{2n}$ and real eigenvalues $\lambda_1, ..., \lambda_{2n}$.

Analysis on Regular Graphs

$P = D^{-1}A = \frac{1}{d}A$ → symmetric $\implies$ orthonormal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_{2n}$ and real eigenvalues $\lambda_1, \ldots, \lambda_{2n}$.

$$\mathbf{x}^{(t)} = P^t \cdot \mathbf{x}^{(0)} = \sum_i \lambda_i^t (\mathbf{v}_i^\intercal \mathbf{x}^{(0)}) \mathbf{v}_i$$

Analysis on Regular Graphs

$P = D^{-1}A = \frac{1}{d}A$ → symmetric $\implies$ orthonormal eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_{2n}$ and real eigenvalues $\lambda_1, ..., \lambda_{2n}$.

$$\mathbf{x}^{(t)} = P^t \cdot \mathbf{x}^{(0)} = \sum_i \lambda_i^t (\mathbf{v}_i^\intercal \mathbf{x}^{(0)})\mathbf{v}_i \xrightarrow{t \to \infty} (\mathbf{v}_1^\intercal \mathbf{x}^{(0)})\mathbf{v}_1$$

Perron-Frobenius Theorem:
$\lambda_1 = 1$, $|\lambda_{i \neq 1}| < 1$

Analysis on Regular Graphs

$P = D^{-1}A = \frac{1}{d}A$ → symmetric $\implies$ orthonormal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_{2n}$ and real eigenvalues $\lambda_1, \ldots, \lambda_{2n}$.

$$\mathbf{x}^{(t)} = P^t \cdot \mathbf{x}^{(0)} = \sum_i \lambda_i^t (\mathbf{v}_i^\intercal \mathbf{x}^{(0)}) \mathbf{v}_i$$

$$\mathbf{v}_1 = \frac{1}{\sqrt{2n}} \mathbf{1}$$

Regular clustered graphs $\implies$ $P\chi = \left(\frac{a-b}{d}\right) \cdot \chi$

Analysis on Regular Graphs

$P = D^{-1}A = \frac{1}{d}A$ → symmetric $\implies$ orthonormal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_{2n}$ and real eigenvalues $\lambda_1, \ldots, \lambda_{2n}$.

$$\mathbf{x}^{(t)} = P^t \cdot \mathbf{x}^{(0)} = \sum_i \lambda_i^t (\mathbf{v}_i^\intercal \mathbf{x}^{(0)}) \mathbf{v}_i$$

$$\mathbf{v}_1 = \frac{1}{\sqrt{2n}} \mathbf{1}$$

Regular clustered graphs $\implies$ $P\chi = \left(\frac{a-b}{d}\right) \cdot \chi$

$$\frac{1}{d}\left(\begin{array}{c|c} \cdots\cdots\cdots & \cdots\cdots\cdots \\ \cdots a \text{ “1”s} \cdots & \cdots b \text{ “1”s} \cdots \\ \cdots\cdots\cdots & \cdots\cdots\cdots \\ \hline \cdots\cdots\cdots & \cdots\cdots\cdots \\ \cdots b \text{ “1”s} \cdots & \cdots a \text{ “1”s} \cdots \\ \cdots\cdots\cdots & \cdots\cdots\cdots \end{array}\right) \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ -1 \end{pmatrix} = \frac{a-b}{d} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ -1 \end{pmatrix}$$

Analysis on Regular Graphs

$P = D^{-1}A = \frac{1}{d}A$ → symmetric $\implies$ orthonormal eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_{2n}$ and real eigenvalues $\lambda_1, ..., \lambda_{2n}$.

$$\mathbf{x}^{(t)} = P^t \cdot \mathbf{x}^{(0)} = \sum_i \lambda_i^t (\mathbf{v}_i^\intercal \mathbf{x}^{(0)}) \mathbf{v}_i$$

$$\mathbf{v}_1 = \frac{1}{\sqrt{2n}} \mathbf{1}$$

Regular clustered graphs $\implies$ $P\chi = \left(\frac{a-b}{d}\right) \cdot \chi$

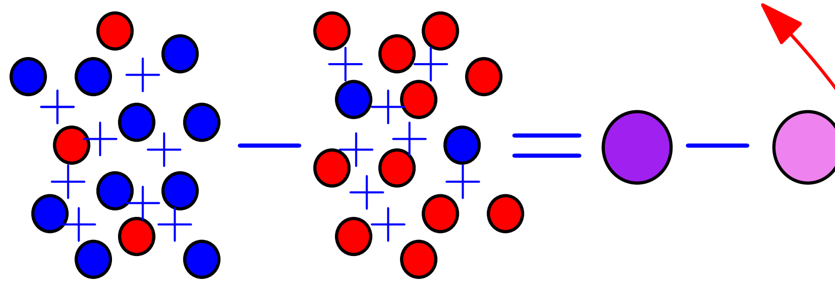
If $\lambda < \frac{a-b}{d} = \lambda_2$ then

$$\mathbf{x}^{(t+1)} = \frac{1}{2n}(\mathbf{1}^\intercal \mathbf{x}^{(0)})\mathbf{1} + \lambda_2^t \frac{1}{2n}(\chi^\intercal \mathbf{x}^{(0)})\chi + \mathbf{e}^{(t)}$$

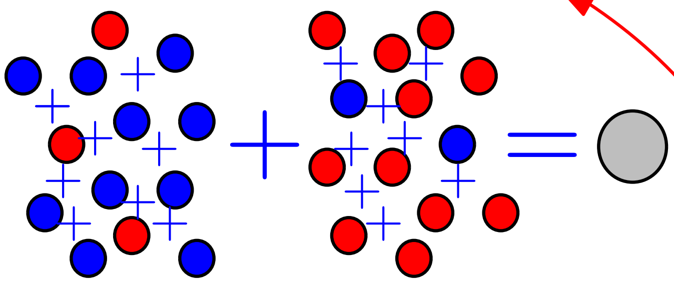
with $\|\mathbf{e}^{(t)}\| = \left\| \sum_{i=3}^{2n} \lambda_i^t (\mathbf{v}_i^\intercal \mathbf{x}^{(0)}) \mathbf{v}_i \right\| \leq \lambda^t \|\mathbf{x}^{(0)}\| \leq \lambda^t \sqrt{2n}$

Analysis on Regular Graphs

$\frac{1}{2}\left(\frac{1}{n}\sum_{u\in V_1}\mathbf{x}^{(0)}(u) - \frac{1}{n}\sum_{u\in V_2}\mathbf{x}^{(0)}(u)\right)$



$\frac{1}{2n}\sum_{u\in V}\mathbf{x}^{(0)}(u)$



If $\lambda < \frac{a-b}{d} = \lambda_2$ then

$$\mathbf{x}^{(t+1)} = \frac{1}{2n}(\mathbf{1}^\intercal\mathbf{x}^{(0)})\mathbf{1} + \lambda_2^t\frac{1}{2n}(\chi^\intercal\mathbf{x}^{(0)})\chi + \mathbf{e}^{(t)}$$

with $\|\mathbf{e}^{(t)}\| = \left\|\sum_{i=3}^{2n}\lambda_i^t(\mathbf{v}_i^\intercal\mathbf{x}^{(0)})\mathbf{v}_i\right\| \le \lambda^t\|\mathbf{x}^{(0)}\| \le \lambda^t\sqrt{2n}$

Analysis on Regular Graphs

If $\boxed{\lambda(1+\delta) < \frac{a-b}{d} = \lambda_2}$ then

$$\mathbf{x}^{(t)} = \frac{1}{2n}(\mathbf{1}^\intercal \mathbf{x}^{(0)})\mathbf{1} + \lambda_2^t \frac{1}{2n}(\chi^\intercal \mathbf{x}^{(0)})\chi + \mathbf{e}^{(t)}$$

with $\|\mathbf{e}^{(t)}\| \leq \lambda^t \sqrt{2n}$

Analysis on Regular Graphs

If $\boxed{\lambda(1+\delta) < \frac{a-b}{d} = \lambda_2}$ then

$$\mathbf{x}^{(t)} = \frac{1}{2n}(\mathbf{1}^\intercal \mathbf{x}^{(0)})\mathbf{1} + \lambda_2^t \frac{1}{2n}(\chi^\intercal \mathbf{x}^{(0)})\chi + \mathbf{e}^{(t)}$$

with $\|\mathbf{e}^{(t)}\| \leq \lambda^t \sqrt{2n}$

$$\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)} = (\chi^\intercal \mathbf{x}^{(0)})\lambda_2^{t-1}(\lambda_2 - 1)\chi + \underbrace{\mathbf{e}^{(t)} - \mathbf{e}^{(t-1)}}_{\ll \lambda_2^{t-1} \text{ if } t = \Omega(\log n)}$$

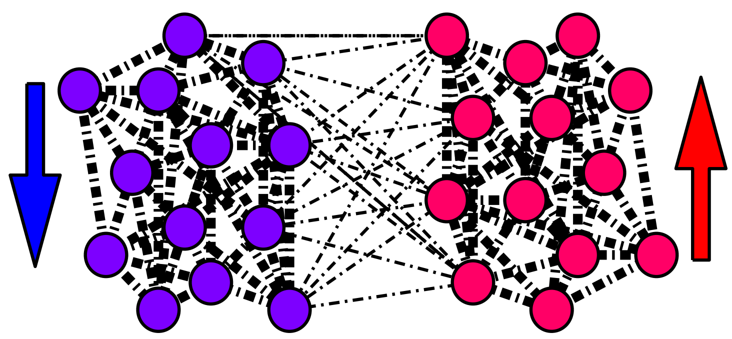
Analysis on Regular Graphs

If $\lambda(1+\delta) < \frac{a-b}{d} = \lambda_2$ then

$$\mathbf{x}^{(t)} = \frac{1}{2n}(\mathbf{1}^\intercal \mathbf{x}^{(0)})\mathbf{1} + \lambda_2^t \frac{1}{2n}(\chi^\intercal \mathbf{x}^{(0)})\chi + \mathbf{e}^{(t)}$$

with $\|\mathbf{e}^{(t)}\| \leq \lambda^t \sqrt{2n}$

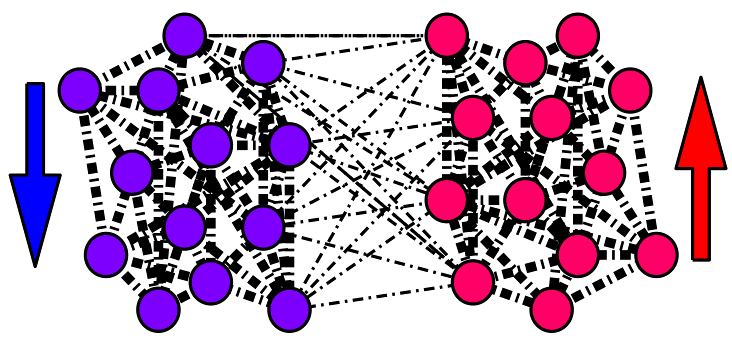
$$\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)} = (\chi^\intercal \mathbf{x}^{(0)})\lambda_2^{t-1}(\lambda_2 - 1)\chi + \underbrace{\mathbf{e}^{(t)} - \mathbf{e}^{(t-1)}}_{\ll \lambda_2^{t-1} \text{ if } t=\Omega(\log n)}$$



$$\text{sign}(\mathbf{x}^{(t)}(u) - \mathbf{x}^{(t-1)}(u)) = \text{sign}(\chi(u)) \text{ or } -\text{sign}(\chi(u))$$

Analysis on Regular Graphs

Corollary.
RSBM is $(2n, d, b)$-clust. regular and $\lambda = \mathcal{O}(\frac{1}{\sqrt{d}}) \ll \frac{a-b}{d}$ by *random degree* k *lifts* [Friedman & Kohler]
$\implies$ Strong reconstruction in $\log n$ w.h.p.



$$\mathrm{sign}(\mathbf{x}^{(t)}(u) - \mathbf{x}^{(t-1)}(u)) = \mathrm{sign}(\chi(u)) \text{ or } -\mathrm{sign}(\chi(u))$$

More Communities

(k,n,d,b)**-clustered Regular Graph.** A graph $G = (\dot{\bigcup}_{i=1}^{k} V_i, E)$ s.t.

- $|V_1| = \cdots = |V_k|$,
- every node has degree $d = a + (k-1)b$
- each $v \in V_i$ has b neighbors in V_j for $j \neq i$.

$\frac{a-b}{d}$ eigenval. with $\mathbf{v}_2, ..., \mathbf{v}_k$ eigenvec. s.t. constant on each V_i and $\mathbf{1}^\intercal \mathbf{v}_i = 0$.

More Communities

(k,n,d,b)**-clustered Regular Graph.** A graph $G = (\dot{\bigcup}_{i=1}^{k} V_i, E)$ s.t.

- $|V_1| = \cdots = |V_k|$,
- every node has degree $d = a + (k-1)b$
- each $v \in V_i$ has b neighbors in V_j for $j \neq i$.

$\frac{a-b}{d}$ eigenval. with $\mathbf{v}_2, ..., \mathbf{v}_k$ eigenvec. s.t. constant on each V_i and $\mathbf{1}^\intercal \mathbf{v}_i = 0$.

$(\frac{a-b}{d} = \lambda_2 = ... = \lambda_k)$

Thm. If $\frac{a-b}{d} > \lambda(1+\delta)$ with $\lambda = \max\{\lambda_{k+1}, |\lambda_{kn}|\}$, then $\Theta(\log n)$ parallel run of averaging gives strong reconstruction in $\mathcal{O}(\log n)$ rounds.

Future Work

Non-regular SBM.
How much “weak” with many communities?

Future Work

Non-regular SBM.
How much "weak" with many communities?

Sparisification.
At each round, pick an edge u.a.r. (*population protocols*):
those two nodes averages their values.
Simulations. Does not (seem to) work for $a - b \ll \log n$.
Analysis. Should work for $a - b \gg \log n$.

Future Work

Non-regular SBM.
How much "weak" with many communities?

Sparisification.
At each round, pick an edge u.a.r. (*population protocols*):
those two nodes averages their values.
Simulations. Does not (seem to) work for $a - b \ll \log n$.
Analysis. Should work for $a - b \gg \log n$.

Planted Clique.
$G_{n,p} \cup$ "clique of $\sqrt{n}(1+\delta)$ nodes":
Does averaging identify the clique?

Thank You!