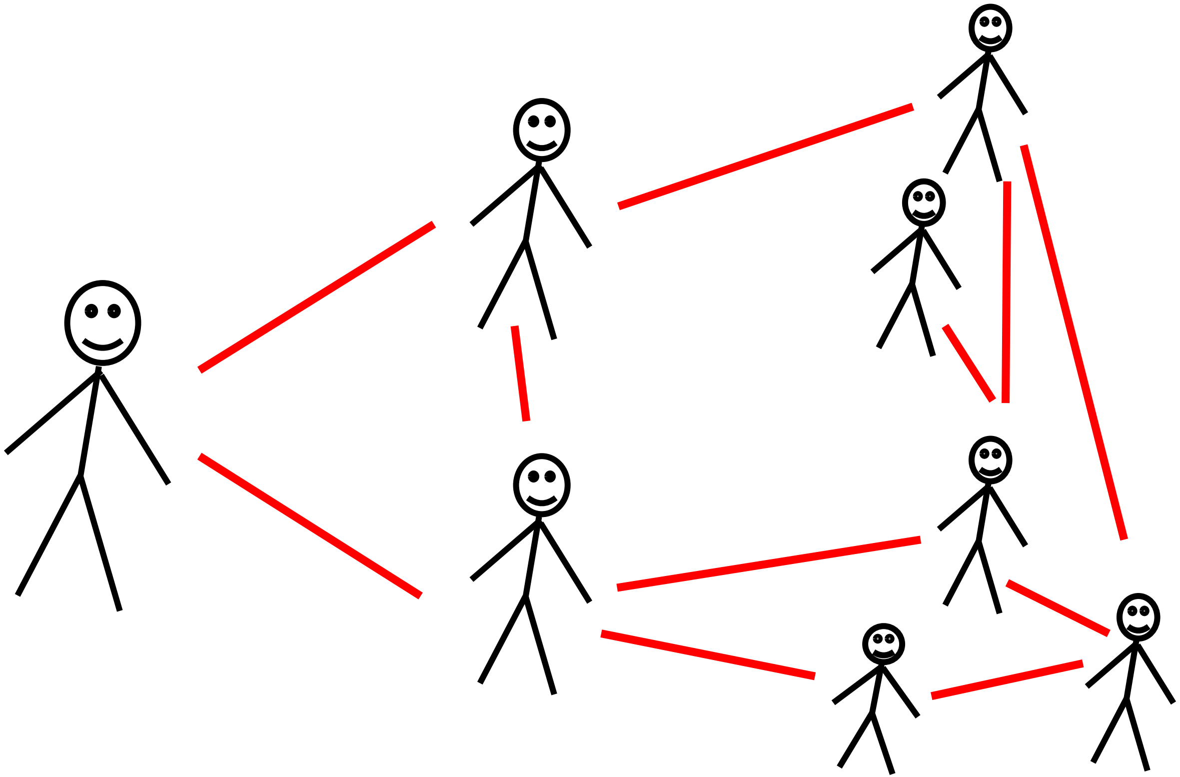
Naive Bayesian Learning in Social Networks

Jerry Anunrojwong (Harvard)
joint with Nat Sothanaphan (MIT)

EC'18

Social Learning

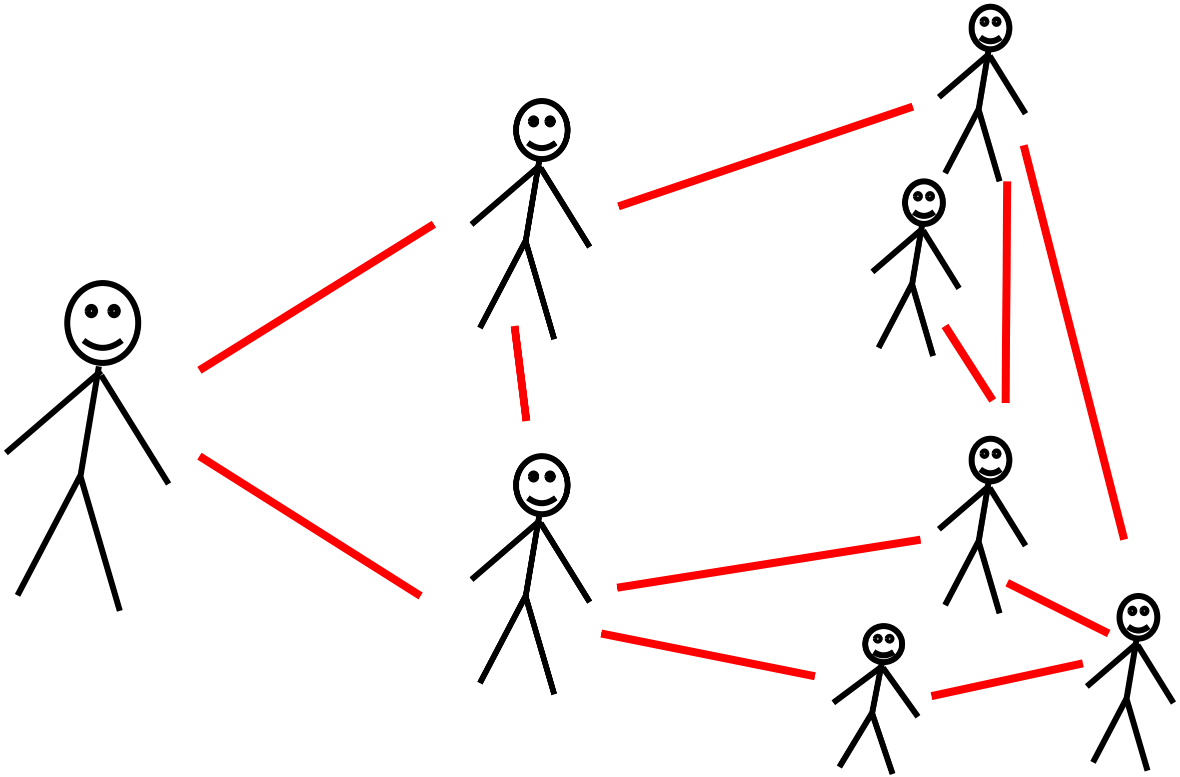


state of the world unknown to the agents

Rule: can only talk to your neighbors

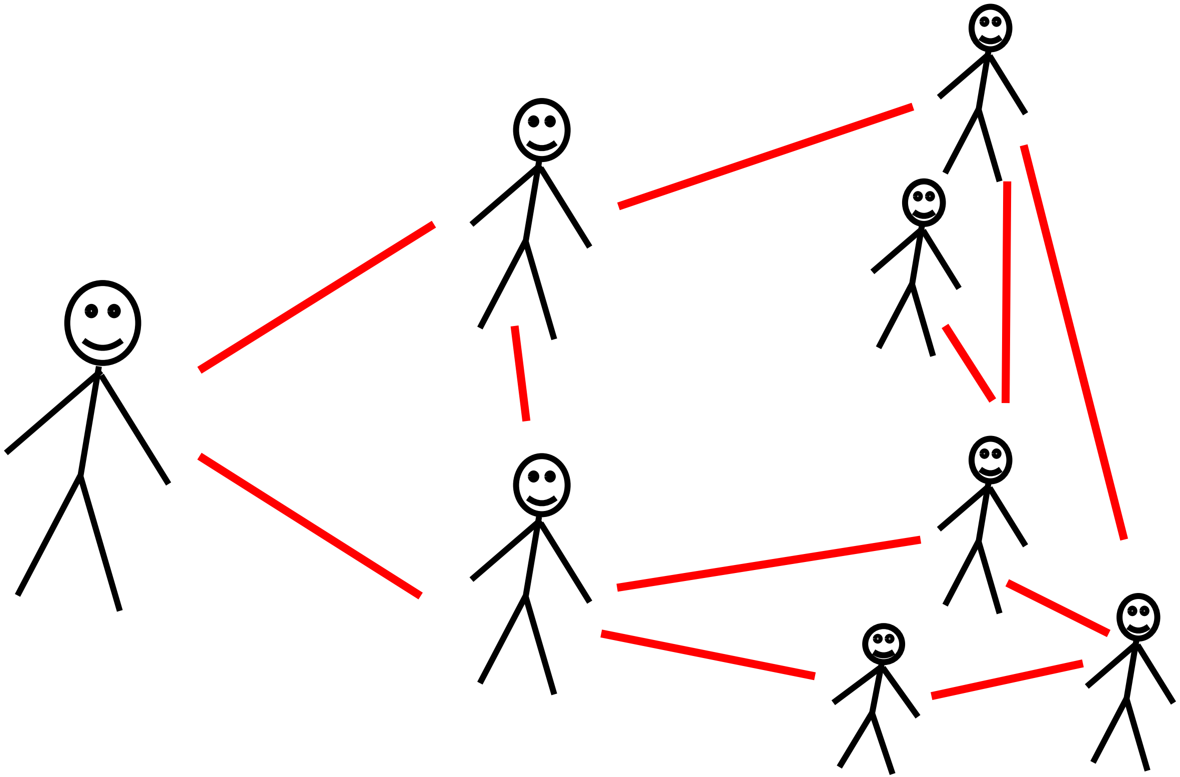
Prior works:
Bayesian Learning
Naive Learning

Bayesian Learning (Rational)



Beliefs are distributions. Perfectly rational and Bayesian.

Bayesian Learning (Rational)

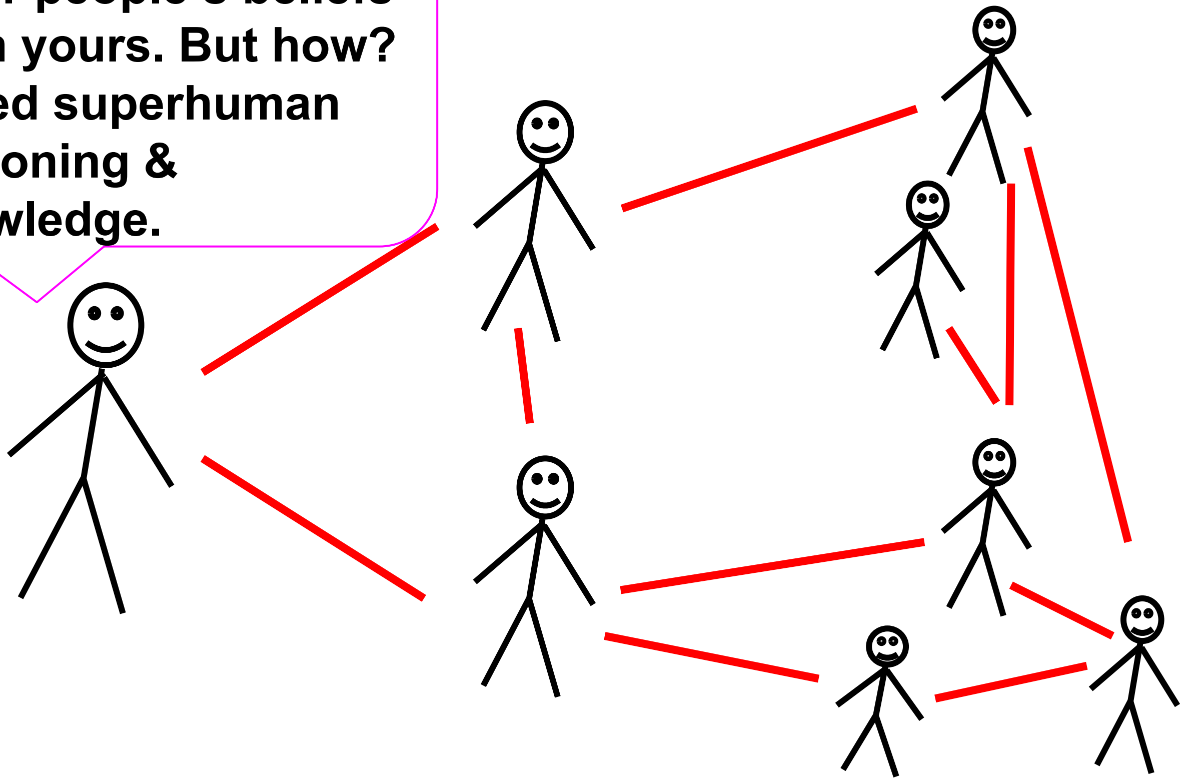


Beliefs are distributions. Perfectly rational and Bayesian.

Weigh confidence in beliefs

Bayesian Learning (Rational)

I need to subtract other people's beliefs from yours. But how? I need superhuman reasoning & knowledge.



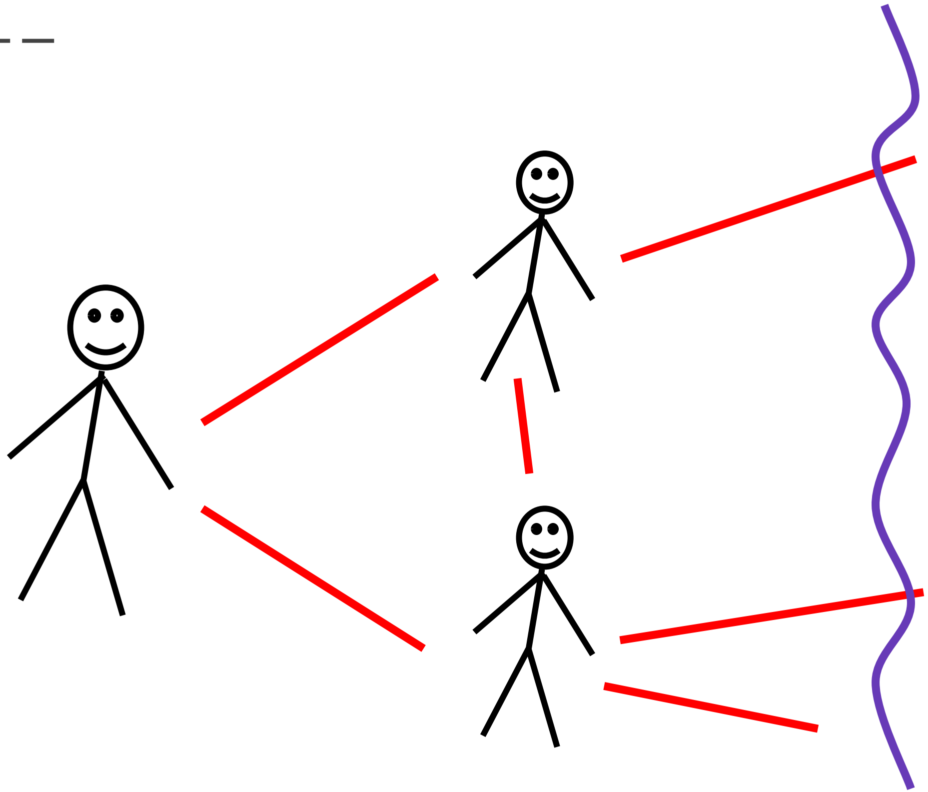
Beliefs are distributions. Perfectly rational and Bayesian.

Weigh confidence in beliefs

Do very sophisticated Bayesian reasoning

Network structure is common knowledge

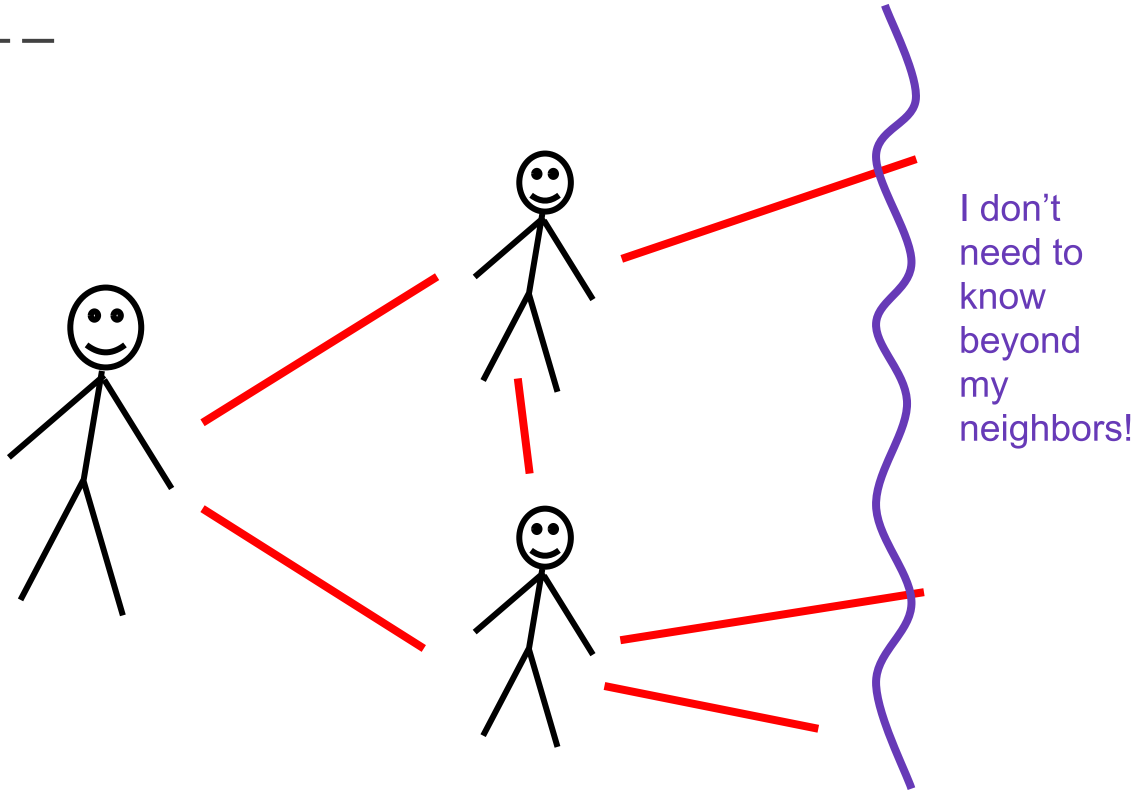
Naive Learning (DeGroot)



I don't
need to
know
beyond
my
neighbors!

Beliefs are scalars. Update beliefs by taking (weighted) average of neighbors' beliefs.

Naive Learning (DeGroot)

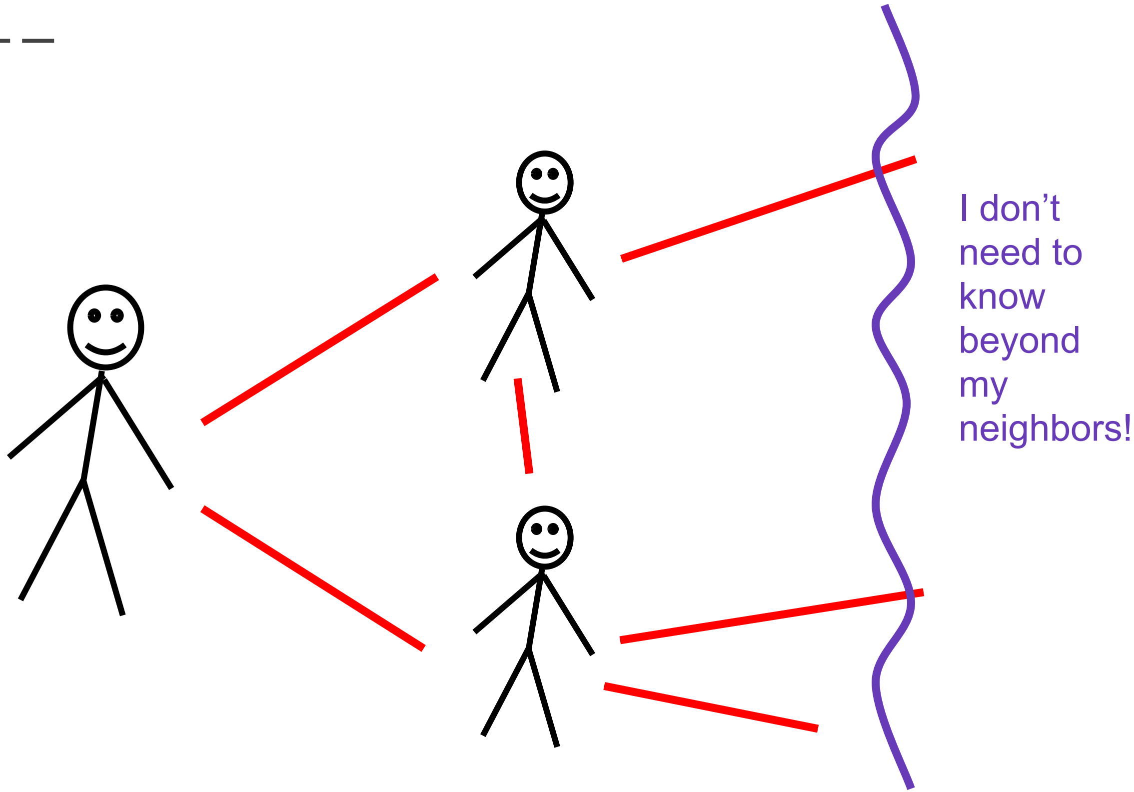


Beliefs are scalars. Update beliefs by taking (weighted) average of neighbors’ beliefs.

Simple and intuitive belief update rule

Only need to know neighbors

Naive Learning (DeGroot)



Beliefs are scalars. Update beliefs by taking (weighted) average of neighbors’ beliefs.

No notion of confidence in beliefs

Simple and intuitive belief update rule

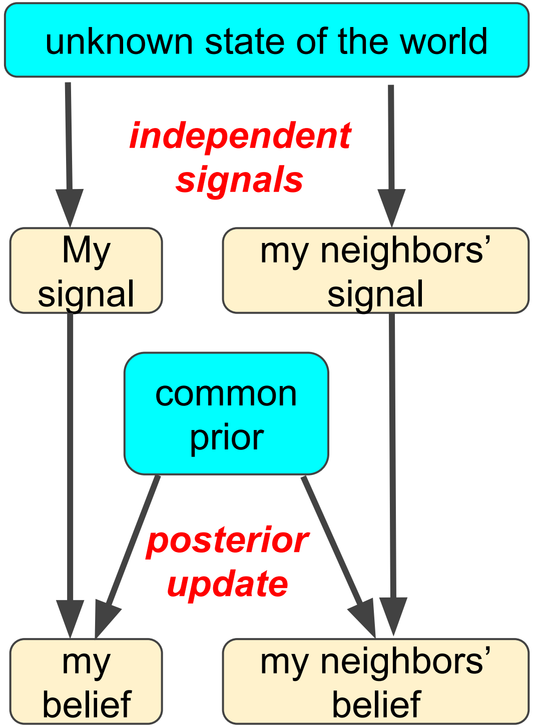
Only need to know neighbors

Question: How can we combine the pros of naive and Bayesian learning?

Naive Bayesian Learning

Weigh confidence in beliefs	**Bayesian**	Beliefs are distributions. Agents use Bayes' rule.
Simple and intuitive belief update rule	**Naive**	Agents treat neighbors as independent.
Only need to know neighbors	**Naive**	Belief update rule only depends on neighbors.

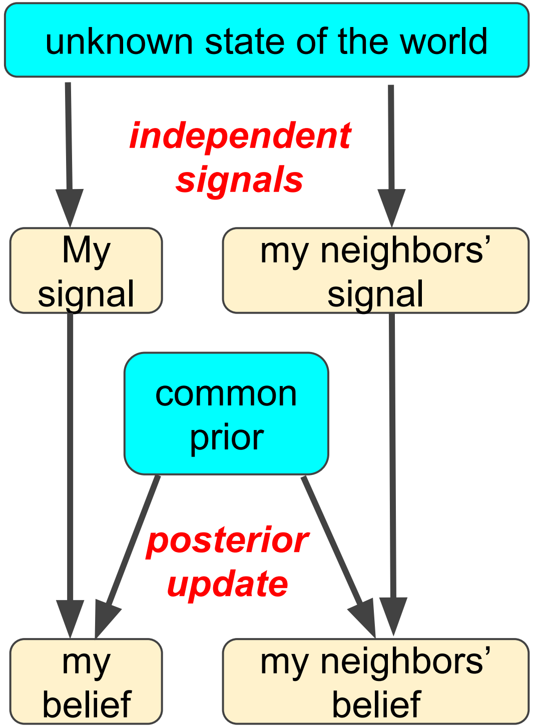
Naive Bayesian Learning (Our paper)



My mental model

Beliefs are distributions. Update beliefs by Bayes rule, *assuming naively that neighbors are independent information sources.*

Naive Bayesian Learning (Our paper)



My mental model

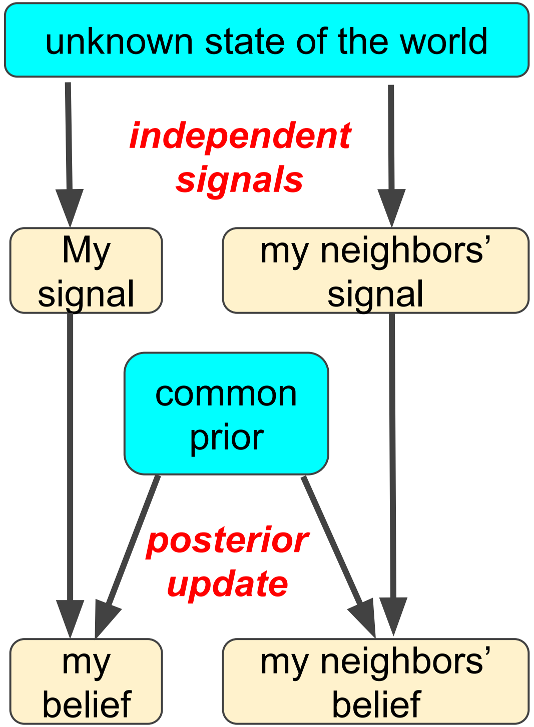
Each time step:

I have access to
my and my neighbors' **beliefs**

My update rule

Beliefs are distributions. Update beliefs by Bayes rule, *assuming naively that neighbors are independent information sources.*

Naive Bayesian Learning (Our paper)



My mental model

Each time step:

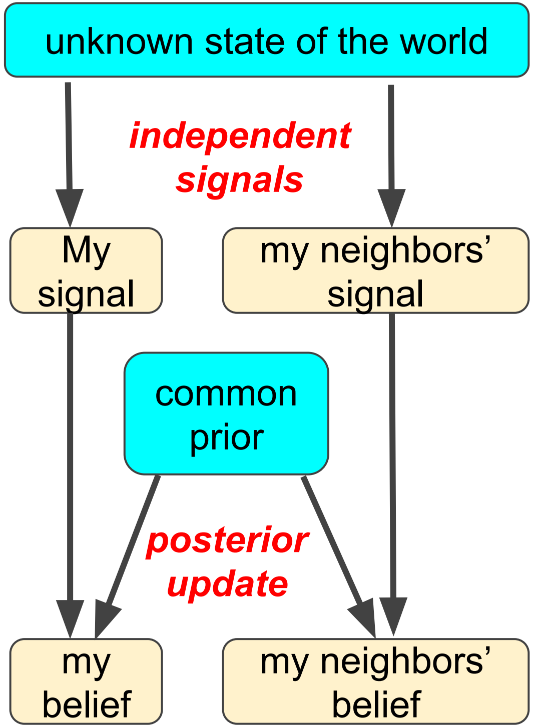
I have access to
my and my neighbors' **beliefs**

I infer my and my neighbors' signals **assuming their beliefs arise from my mental model**

My update rule

Beliefs are distributions. Update beliefs by Bayes rule, *assuming naively that neighbors are independent information sources.*

Naive Bayesian Learning (Our paper)



My mental model

Each time step:

I have access to my and my neighbors' **beliefs**

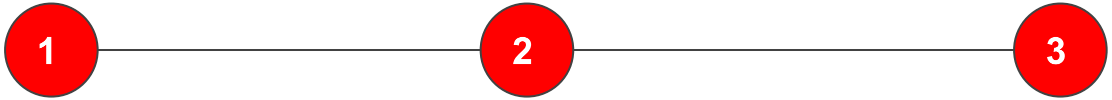
I infer my and my neighbors' signals **assuming their beliefs arise from my mental model**

I update my beliefs from common prior by **conditioning on my and my neighbors' inferred signals**

My update rule

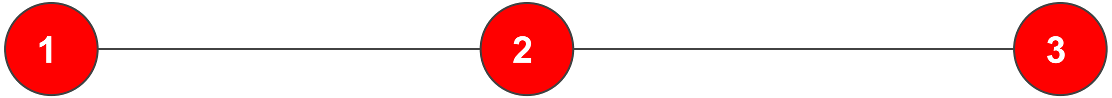
Beliefs are distributions. Update beliefs by Bayes rule, *assuming naively that neighbors are independent information sources.*

Naive Bayesian Update Rule: Example



belief t=0	$p(\theta \mid X_1)$	$p(\theta \mid X_2)$	$p(\theta \mid X_3)$
inferred signals	$\{X_1\}$	$\{X_2\}$	$\{X_3\}$
belief t=1	$p(\theta \mid X_1, X_2)$	$p(\theta \mid X_1, X_2, X_3)$	$p(\theta \mid X_2, X_3)$

Naive Bayesian Update Rule: Example

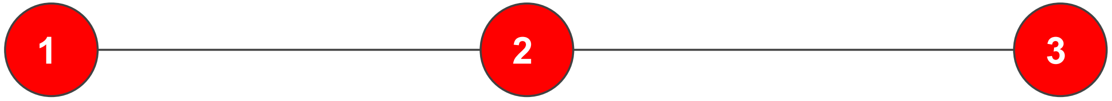


belief t=1	$p(\theta \| X_1, X_2)$	$p(\theta \| X_1, X_2, X_3)$	$p(\theta \| X_2, X_3)$
inferred signals	$\{X_1, X_2\}$	$\{X_1, X_2, X_3\}$	$\{X_2, X_3\}$
belief t=2	$p(\theta \| X_1, X_2, X_1, X_2, X_3)$	$p(\theta \| X_1, X_2, X_1, X_2, X_3, X_2, X_3)$	$p(\theta \| X_1, X_2, X_3, X_2, X_3)$

Naive Bayesian Update Rule: Example

Copies of signals "flow"

Mental model assumes beliefs are "fresh"



	1	2	3
belief t=1	$p(\theta \vert X_1, X_2)$	$p(\theta \vert X_1, X_2, X_3)$	$p(\theta \vert X_2, X_3)$
inferred signals	$\{X_1, X_2\}$	$\{X_1, X_2, X_3\}$	$\{X_2, X_3\}$
belief t=2	$p(\theta \vert X_1, X_2, X_1, X_2, X_3)$	$p(\theta \vert X_1, X_2, X_1, X_2, X_3, X_2, X_3)$	$p(\theta \vert X_1, X_2, X_3, X_2, X_3)$

Main Result (Informal)

we **analytically characterize the consensus** and the formula for the consensus says ...

influence on consensus = **centrally located** + **confident beliefs**

NAIVE LEARNING

BAYESIAN LEARNING

Main Result (Informal)

we **analytically characterize the consensus** and the formula for the consensus says ...

influence on consensus = **centrally located** + **confident beliefs**

NAIVE LEARNING

BAYESIAN LEARNING

---

eigenvector centrality

eigenvector of adjacency matrix

$$A\mathbf{v} = r\mathbf{v}$$

Main Result (Informal)

we **analytically characterize the consensus** and the formula for the consensus says ...

influence on consensus = **centrally located** + **confident beliefs**

NAIVE LEARNING

BAYESIAN LEARNING

---

eigenvector centrality

eigenvector of adjacency matrix

an agent is central if it connects to other central agents

$$\mathbf{A}\mathbf{v} = r\mathbf{v} \Rightarrow rv_i = \sum_{j \in N(i)} v_j$$

Main Result (Informal)

we **analytically characterize the consensus** and the formula for the consensus says ...

influence on consensus = **centrally located** + **confident beliefs**

NAIVE LEARNING

BAYESIAN LEARNING

---

eigenvector centrality

eigenvector of adjacency matrix

an agent is central if it connects to other central agents

$$\mathbf{A}\mathbf{v} = r\mathbf{v} \Rightarrow r v_i = \sum_{j \in N(i)} v_j$$

also appear in DeGroot learning but from different dynamics

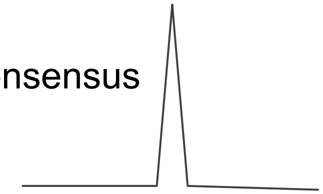
Main Result (Formal)

DEFINITION
weighted log-likelihood function

$$\ell(\theta) = \sum_{i=1}^{n} v_i \log\left(\frac{f_i(\theta)}{f_*(\theta)}\right)$$

for each state θ

consensus



θ_{max}

<u>Theorem</u> Every agent’s belief converges to the point distribution at maximizer of $\ell(\theta)$.

Main Result (Formal)

f_i agent i's initial belief

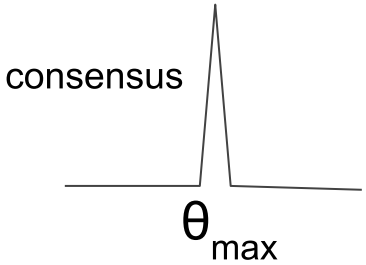
f_* common prior

DEFINITION
weighted log-likelihood function

$$\ell(\theta) = \sum_{i=1}^{n} v_i \log\left(\frac{f_i(\theta)}{f_*(\theta)}\right)$$

for each state θ

"confidence of beliefs at θ"
how much agent i believes in θ
compared to the prior baseline



<u>Theorem</u> Every agent's belief converges to the point distribution at maximizer of ℓ(θ).

f_i agent i’s initial belief

f_* common prior

Main Result (Formal)

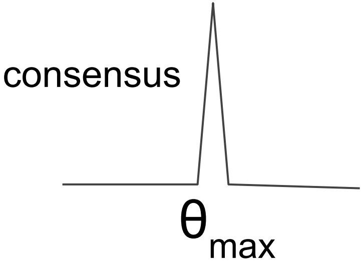
DEFINITION
weighted log-likelihood function

$$\ell(\theta) = \sum_{i=1}^{n} v_i \log\left(\frac{f_i(\theta)}{f_*(\theta)}\right)$$

for each state θ

centrality-weighted average of

“confidence of beliefs at θ”
how much agent i believes in θ
compared to the prior baseline



<u>Theorem</u> Every agent’s belief converges to the point distribution at maximizer of ℓ(θ).

Understanding Main Result Intuitively

- **agents take a lot of signals as independent**
  → beliefs converge to a point

Understanding Main Result Intuitively

- agents take a lot of signals as independent
  → beliefs converge to a point
- initial beliefs come from independent signals
  → “confident beliefs” = “informative signals”

Example: Gaussian Beliefs

agent i's initial belief

$$\mathcal{N}(\mu_i, 1/\tau_i)$$

interpretation: scalar belief μ_i with confidence τ_i

Example: Gaussian Beliefs

agent i's initial belief

$$\mathcal{N}(\mu_i, 1/\tau_i)$$

interpretation: scalar belief μ_i with confidence τ_i

a scenario: at the beginning, agent i receives signal
$\mu_i = \theta^* + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, 1/\tau_i)$ independent

Example: Gaussian Beliefs

agent i's initial belief

$$\mathcal{N}(\mu_i, 1/\tau_i)$$

interpretation: scalar belief μ_i with confidence τ_i

a scenario: at the beginning, agent i receives signal
$\mu_i = \theta^* + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, 1/\tau_i)$ independent

consensus

$$\theta_{\max} = \sum_{i=1}^{n} c_i \mu_i$$

Example: Gaussian Beliefs

agent i's initial belief

$$\mathcal{N}(\mu_i, 1/\tau_i)$$

interpretation: scalar belief μ_i with confidence τ_i

a scenario: at the beginning, agent i receives signal
$\mu_i = \theta^* + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, 1/\tau_i)$ independent

consensus

$$\theta_{\max} = \sum_{i=1}^{n} c_i \mu_i$$

agent i's influence

$$c_i = \frac{v_i \tau_i}{\sum_{j=1}^{n} v_j \tau_j} \propto v_i \tau_i$$

influence on consensus = **centrally located** + **informative signals**

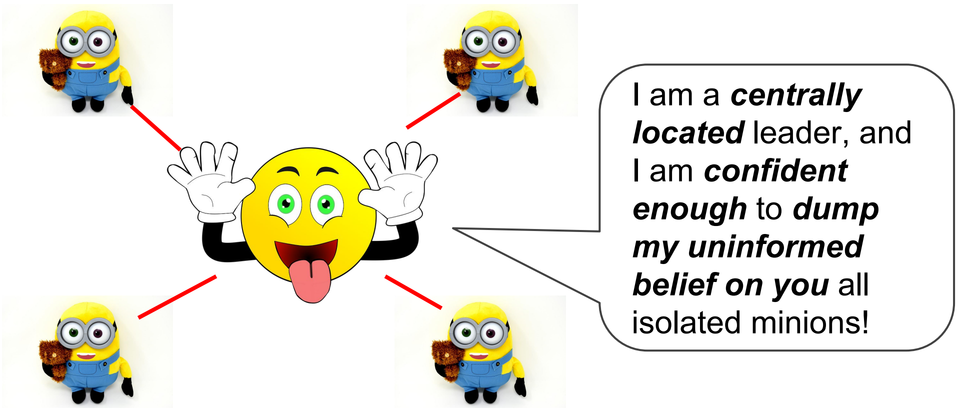
Policy Implication I: how to seed opinion leaders

Learning quality = precision of consensus, as a random variable!

$$\frac{\partial(\text{learning quality})}{\partial(\text{signal precision})} > 0$$ unless agent is central but poorly informed

If social planners want to seed opinion leaders, they must make those leaders well informed.

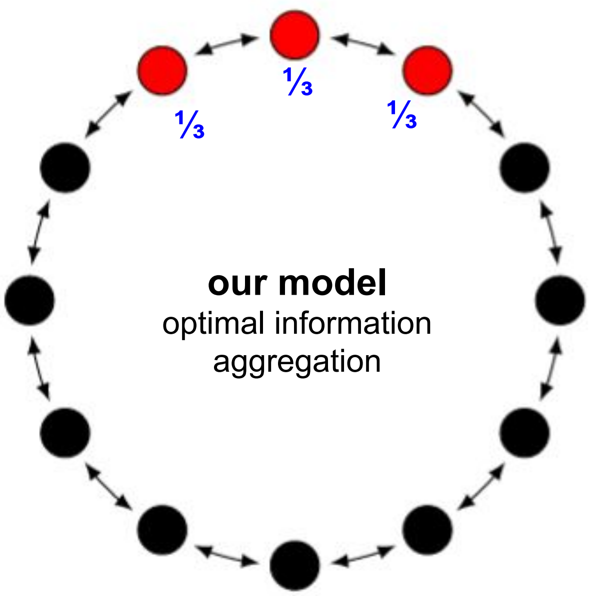
ELSE you get this

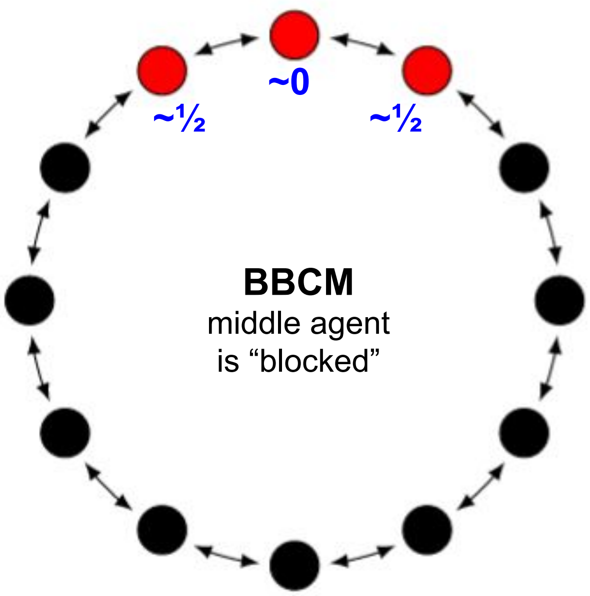


Policy Implication II: how to solve clustered seeding

Key point: their model has no notion of "confidence in beliefs"

Information loss from *clustered seeding* occurs in their model but not ours.





Conclusion

- We propose a model that combines
  the pros of naive and Bayesian learning.
- consensus = maximizer of the weighted log-likelihood function
- centrally located + confident beliefs = influence on consensus
- Two policy implications:
  how to seed opinion leaders + clustered seeding

Gaussian Beliefs: Quality of Learning

$$\theta_{\max} \sim \mathcal{N}(\theta^*, 1/Q) \qquad Q = \frac{\left(\sum_i v_i \tau_i\right)^2}{\sum_i v_i^2 \tau_i}$$

consensus is a random variable
precision Q captures learning quality

Comparative statics

$$\boxed{\frac{\partial Q}{\partial \tau_k} > 0 \ \text{ unless } v_k \text{ is large and } \tau_k \text{ is small}}$$

POLICY IMPLICATION

If social planners want to seed opinion leaders, they must make those leaders well informed.

ELSE you get this