Some Models of Information Aggregation and Consensus in Networks

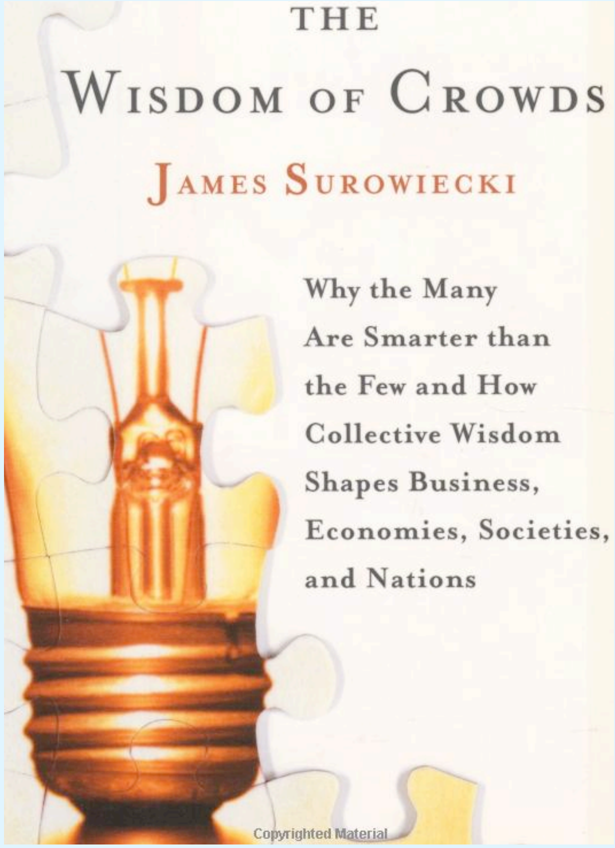
John N. Tsitsiklis
MIT

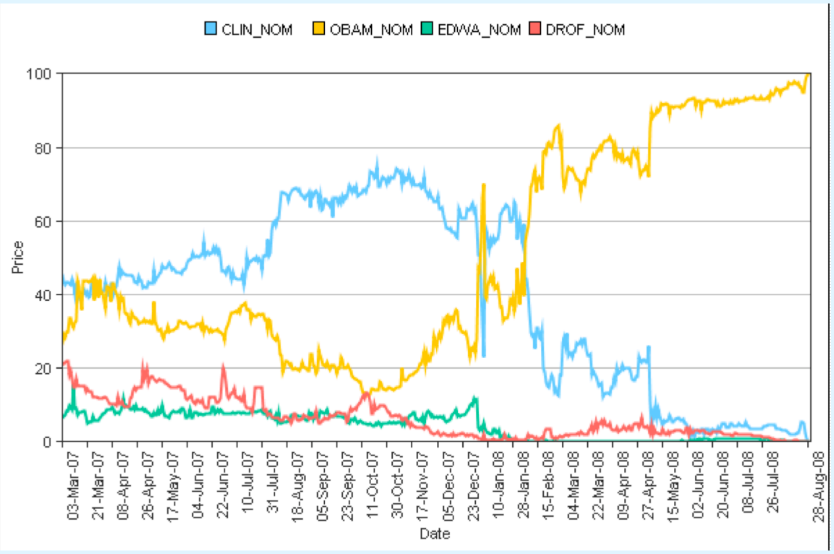
ACCESS, KTH
January 2011

Overview

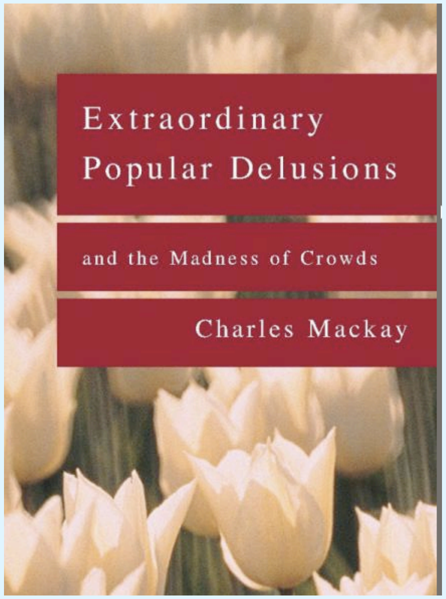
- Information propagation/aggregation in networks
  - Engineered versus social networks
  - Bayesian versus naive updates
- Review a few key models and results

The Wisdom of Crowds





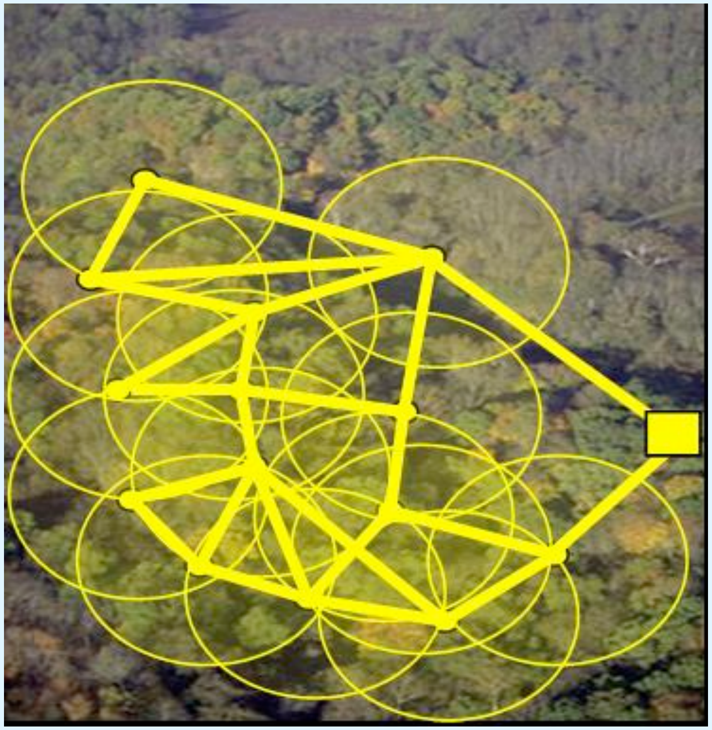
Wisdom or Madness of Crowds?



- Charles Mackay (London, 1841): Extraordinary Popular Delusions and the Madness of Crowds
  - "Men, it has been well said, think in herds; it will be seen that they go mad in herds,..."

Sensor and other Engineered Networks

- Fusion of available information
  - we get to design the nodes' behavior



The Basic Setup

- Each node i endowed with private information X_i
  - Nodes form opinions, make decisions
  - Nodes observe opinions/decisions or receive messages
  - Nodes update opinions/decisions
  - Everyone wants a "good" decision; common objective

The Basic Setup

- Each node i endowed with private information X_i
  - Nodes form opinions, make decisions
  - Nodes observe opinions/decisions or receive messages
  - Nodes update opinions/decisions
  - Everyone wants a "good" decision; common objective
- **Social science:** postulate update mechanism
- **Engineering:** design/optimize update mechanism

The Basic Setup

- Each node i endowed with private information X_i
  - Nodes form opinions, make decisions
  - Nodes observe opinions/decisions or receive messages
  - Nodes update opinions/decisions
  - Everyone wants a "good" decision; common objective
- **Social science:** postulate update mechanism
- **Engineering:** design/optimize update mechanism
- Questions
  - Convergence? To what? How fast? Quality of limit?
  - Does the underlying graph matter? (Tree, acyclic, random,...

Bayesian Models

Where do we eat tonight?

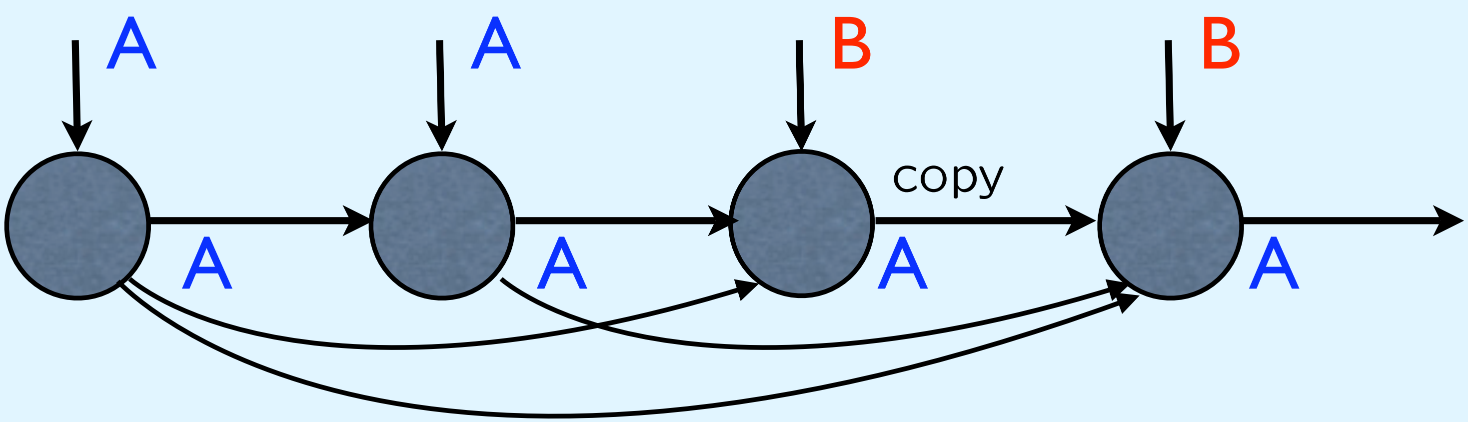
(Bickchandani, Hisrchleifer, Welch, 92; Banerjee, 92)

- Private info points to the better one, with some probability of error

Where do we eat tonight?

(Bickchandani, Hisrchleifer, Welch, 92; Banerjee, 92)

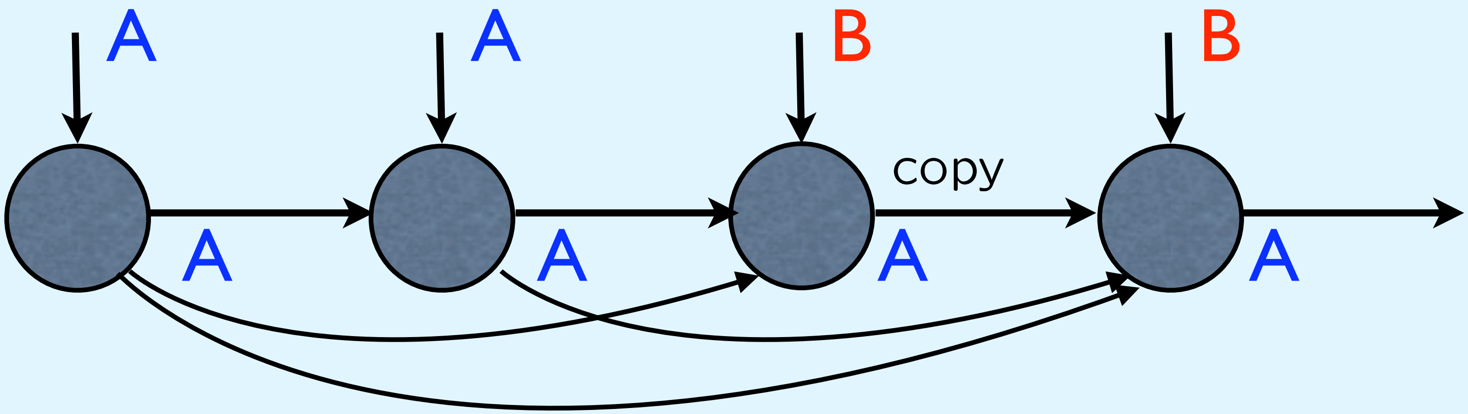
- Private info points to the better one, with some probability of error



Where do we eat tonight?

(Bickchandani, Hisrchleifer, Welch, 92; Banerjee, 92)

- Private info points to the better one, with some probability of error

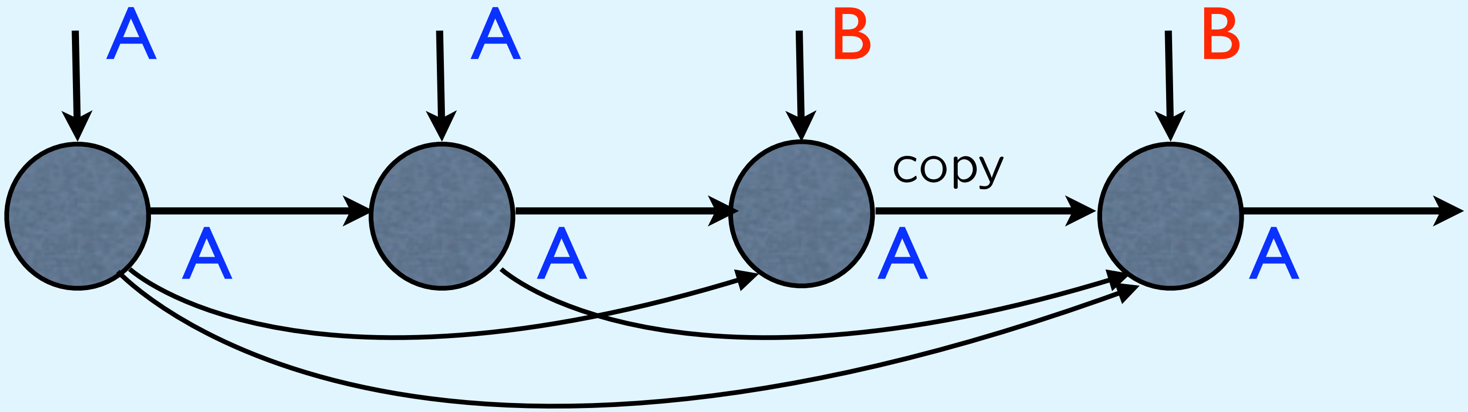


- Information is there

Where do we eat tonight?

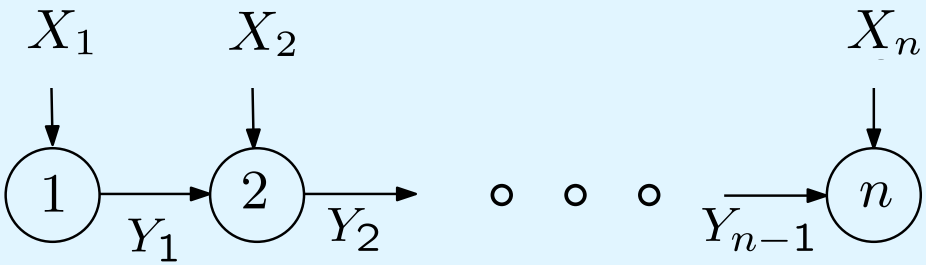
(Bickchandani, Hisrchleifer, Welch, 92; Banerjee, 92)

- Private info points to the better one, with some probability of error



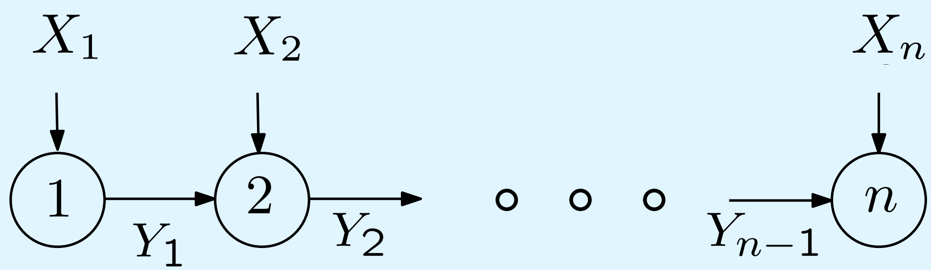
- Information is there but is not used

Tandem networks



- X_i are i.i.d.; $H_0 : X_i \sim \mathbb{P}_0$, $H_1 : X_i \sim \mathbb{P}_1$
- binary message/decision functions γ_i: $Y_i = \gamma_i(X_i, Y_{i-1})$

Tandem networks



- X_i are i.i.d.; $H_0 : X_i \sim \mathbb{P}_0$, $H_1 : X_i \sim \mathbb{P}_1$
- binary message/decision functions γ_i: $Y_i = \gamma_i(X_i, Y_{i-1})$

- "Social network" view: each γ_i is myopic–Bayes optimal
- "Engineered system view": $\{\gamma_i\}$ designed for best end-decision
  - simple sensor network [Cover, 1969]

- $e_n = \mathbb{P}(Y_n \text{ is incorrect}) \to 0$?

The Two Regimes

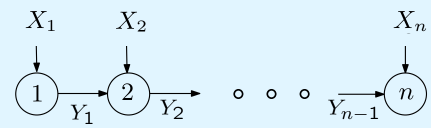
Bounded likelihood ratios (B-LR): $0 < a < \frac{d\,\mathbb{P}_1}{d\,\mathbb{P}_0}(X_i) < b < \infty$

never get compelling evidence

Unbounded likelihood ratios (U-LR): $0 < \frac{d\,\mathbb{P}_1}{d\,\mathbb{P}_0}(X_i) < \infty$

arbitrarily compelling evidence is possible

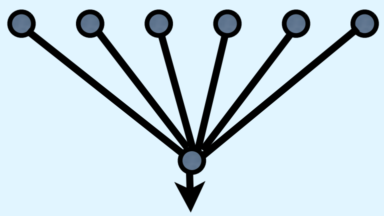
Tandem Network Results



	Social		Engineered
U-LR:	$e_n \to 0$ (Papastavrou & Athans, 92)	$\Longrightarrow$	$e_n \to 0$ (Cover, 69) but slowly (Tay, Win & JNT, 08)
B-LR:	$e_n \not\to 0$	$\Longleftarrow$	$e_n \not\to 0$ (Koplowitz, 75)
B-LR, ternary messages	$e_n \not\to 0$ (Dia & JNT, 09)		$e_n \to 0$ (Koplowitz, 75)

Engineered — Architectural Comparisons

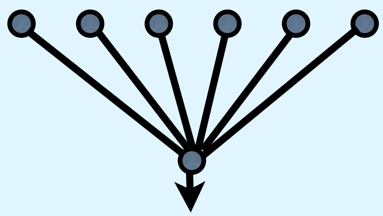
Engineered — Architectural Comparisons



$e_n \sim \exp\{-\lambda n\}$

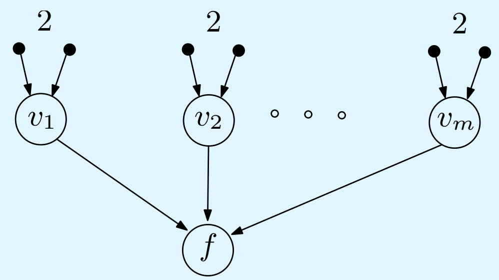
λ is known (JNT, 88)

Engineered — Architectural Comparisons



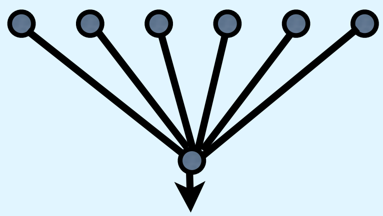
$e_n \sim \exp\{-\lambda n\}$

λ is known (JNT, 88)



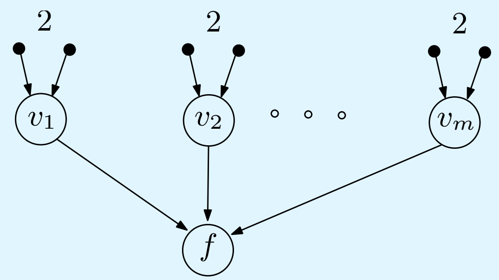
worse error exponent

Engineered — Architectural Comparisons

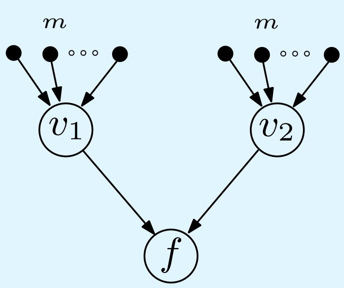


$e_n \sim \exp\{-\lambda n\}$

λ is known (JNT, 88)



worse error exponent

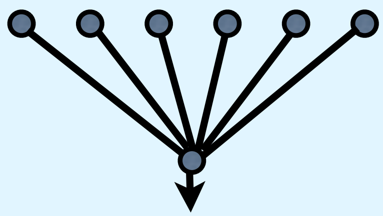


less information

but "same performance"

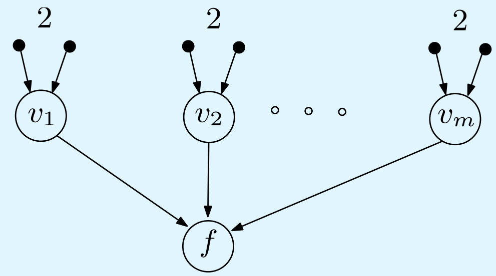
(Tay, Win & JNT, 08)

Engineered — Architectural Comparisons

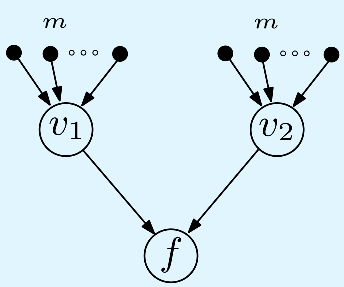


$e_n \sim \exp\{-\lambda n\}$

λ is known (JNT, 88)



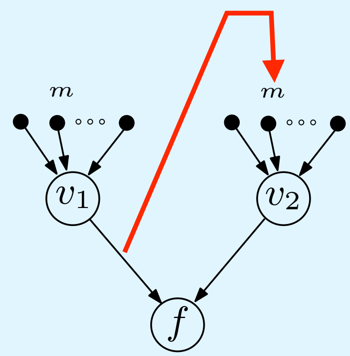
worse error exponent



less information

but "same performance"

(Tay, Win & JNT, 08)



more information

but "same performance"

(Kreidl, Zoumpoulis & JNT, 10)

A Perspective on Bayesian Optimality

- **Engineering**
  - Optimal rules are hard to find
  - Can design "good" schemes (asymptotic, etc.)

- **Social Networks**
  - Is Bayesian updating plausible?

“Naive” Information Aggregation (Consensus and Averaging)

The Setting

- n agents
  - starting values $x_i(0)$
- reach consensus on some x^*, with either:
  - $\min_i x_i(0) \leq x^* \leq \max_i x_i(0)$ (consensus)
  - $x^* = \frac{x_1(0) + \cdots + x_n(0)}{n}$ (averaging)

The Setting

- n agents
  - starting values $x_i(0)$
- reach consensus on some x^*, with either:
  - $\min_i x_i(0) \leq x^* \leq \max_i x_i(0)$ (consensus)
  - $x^* = \dfrac{x_1(0) + \cdots + x_n(0)}{n}$ (averaging)
- simple updates, such as: $x_i := \dfrac{x_i + x_j}{2}$

Social sciences

- Merging of “expert” opinions
- Evolution of public opinion
- Evolution of reputation
- Modeling of jurors
- Language evolution

- interested in modeling, analysis (descriptive theory)
  - ... and narratives

Engineering

- Distributed computation and sensor networks
  - Fusion of individual estimates
  - Distributed Kalman filtering
  - Distributed optimization
  - Distributed reinforcement learning
- Networking
  - Load balancing and resource allocation
  - Clock synchronization
  - Reputation management in ad hoc networks
  - Network monitoring
- Multiagent coordination and control
  - Coverage control
  - Monitoring
  - Creating virtual coordinates for geographic routing
  - Decentralized task assignment
  - Flocking

Distributed optimization

$$\min_x f(x)$$

centralized $\quad x := x - \gamma \nabla f(x) + \text{noise}$

distributed $\quad x^i := x - \gamma \nabla f(x) + \text{noise}_i$

Distributed optimization

	$\min_x f(x)$	$\min_x \sum_{i=1}^n f_i(x)$
centralized	$x := x - \gamma \nabla f(x) + \text{noise}$	$x := x - \gamma \sum_{i=1}^n \nabla f_i(x)$
distributed	$x^i := x - \gamma \nabla f(x) + \text{noise}_i$	$x^i := x^i - \gamma \nabla f_i(x^i)$

Distributed optimization

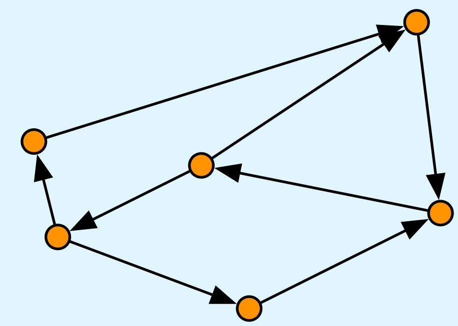
	$\min_x f(x)$	$\min_x \sum_{i=1}^{n} f_i(x)$
centralized	$x := x - \gamma \nabla f(x) + \text{noise}$	$x := x - \gamma \sum_{i=1}^{n} \nabla f_i(x)$
distributed	$x^i := x - \gamma \nabla f(x) + \text{noise}_i$	$x^i := x^i - \gamma \nabla f_i(x^i)$

reconcile updates through a consensus or averaging algorithm

The DeGroot opinion pooling model (1974)

$x_i(t+1) = \sum_j a_{ij}\, x_j(t)$ $\qquad$ $a_{ij} \geq 0, \quad \sum_j a_{ij} = 1$

$x(t+1) = Ax(t)$ $\qquad$ A: stochastic matrix



- Markov chain theory + "mixing conditions"

  $\longrightarrow$ convergence of A^t, to matrix with equal rows

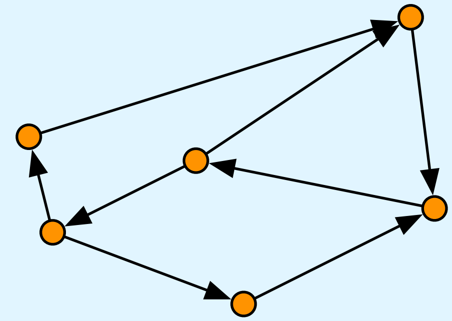
  $\longrightarrow$ convergence of x_i to $\sum_j \pi_j x_j$

  $\longrightarrow$ convergence rate estimates

The DeGroot opinion pooling model (1974)

$x_i(t+1) = \sum_j a_{ij}\, x_j(t)$ $\qquad a_{ij} \geq 0, \quad \sum_j a_{ij} = 1$

$x(t+1) = Ax(t)$ $\qquad A$: stochastic matrix



- Markov chain theory + "mixing conditions"

  $\longrightarrow$ convergence of A^t, to matrix with equal rows

  $\longrightarrow$ convergence of x_i to $\sum_j \pi_j x_j$

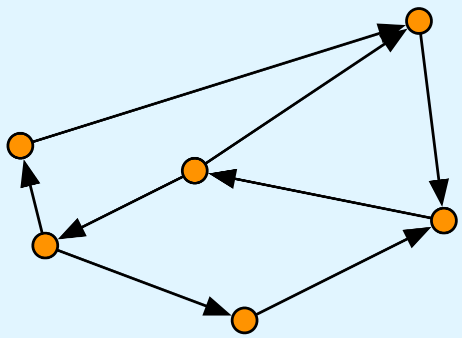
  $\longrightarrow$ convergence rate estimates

- Averaging algorithms
  - A doubly stochastic: $\mathbf{1}'\, A\, x = \mathbf{1}'\, x$, $\quad$ where $\mathbf{1}' = [1\ 1\ \ldots\ 1]$
  - $x_1 + \cdots + x_n$ is conserved
  - convergence to $x^* = \dfrac{x_1(0) + \cdots + x_n(0)}{n}$

Convergence time of consensus algorithms

$x_i(t+1) = \sum_j a_{ij}\, x_j(t)$ $\qquad$ $a_{ij} \geq 0, \quad \sum_j a_{ij} = 1$

$x(t+1) = Ax(t)$ $\qquad$ A: stochastic matrix

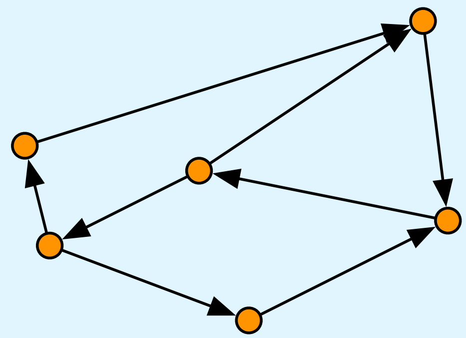


Convergence time (time to get close to "steady-state")

Convergence time of consensus algorithms

$$x_i(t+1) = \sum_j a_{ij}\, x_j(t) \qquad a_{ij} \geq 0, \quad \textstyle\sum_j a_{ij} = 1$$

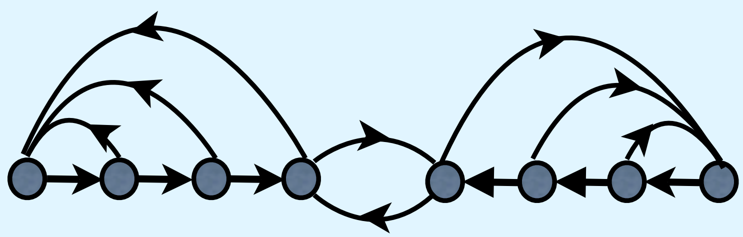
$$x(t+1) = Ax(t) \qquad A\text{: stochastic matrix}$$



Convergence time (time to get close to "steady-state")

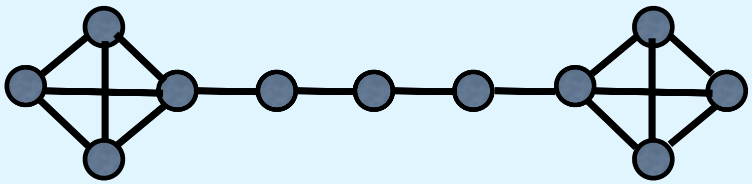
Equal weight to all neighbors

Directed graphs: exponential(n)



Undirected graphs: $O(n^3)$, tight

(Landau and Odlyzko, 1981)



A critique

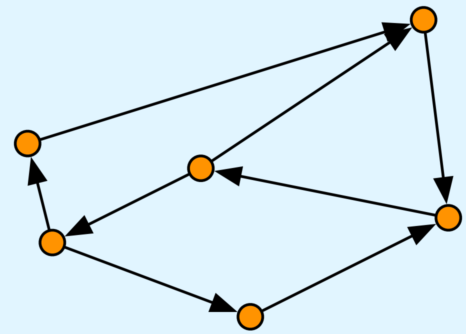
- **Social Networks**
  - "The process that it describes is intuitively appealing." (DeGroot, 74)
  - How plausible is this type of synchronism?

- **Engineering**
  - If the graph is indeed fixed:
    elect a leader, form a spanning tree, accumulate on tree
  - Want simplicity, and robustness w.r.t. changing topologies, failures, etc.
  - Different models and specs?

Time-Varying/Chaotic Environments

- Fairly arbitrary sequence of graphs/matrices $A(t)$: worst-case analysis

$$x_i(t+1) = \sum_j a_{ij}(t) x_j(t)$$



$a_{ij}(t)$: nonzero whenever i receives message from j

Consensus convergence

$$x_i(t+1) = \sum_j a_{ij}(t) x_j\big(t\big)$$

- $a_{ii}(t) > 0$; $\qquad a_{ij}(t) > 0 \implies a_{ij}(t) \geq \alpha > 0$
- "strong connectivity in bounded time":
  over B time steps "communication graph"
  is strongly connected
- Convergence to consensus:
  $\forall\ i: \quad x_i(t) \to x^* =$ convex combination of initial values
  (JNT, Bertsekas, Athans, 86; Jadbabaie et al., 03)
- "convergence time": exponential in n and B
  - even with:
    symmetric graph at each time
    equal weight to each neighbor
    (Cao, Spielman, Morse, 05)

Averaging in Time-Varying Setting

- $x(t+1) = A(t)x(t)$
  - $A(t)$ doubly stochastic, for all t,
  - nonzero $a_{ij}(t) \geq \alpha > 0$
  - $O(n^2/\alpha)$ [Nedic, Olshevsky, Ozdaglar, JNT, 09]

Averaging in Time-Varying Setting

- $x(t+1) = A(t)x(t)$
  - $A(t)$ doubly stochastic, for all t,
  - nonzero $a_{ij}(t) \geq \alpha > 0$
  - $O(n^2/\alpha)$ [Nedic, Olshevsky, Ozdaglar, JNT, 09]

- Improved convergence rate
  - exchange "load" with up to two neighbors at a time
  - can use $\alpha = O(1)$
  - convergence time: $O(n^2)$

Averaging in Time-Varying Setting

- $x(t+1) = A(t)x(t)$
  - $A(t)$ doubly stochastic, for all t,
  - nonzero $a_{ij}(t) \geq \alpha > 0$
  - $O(n^2/\alpha)$ [Nedic, Olshevsky, Ozdaglar, JNT, 09]

- Improved convergence rate
  - exchange "load" with up to two neighbors at a time
  - can use $\alpha = O(1)$
  - convergence time: $O(n^2)$

- Is there a $\Omega(n^2)$ bound to be discovered?

Closing Thoughts

Engineering

- **Bayesian:** near-optimal designs possible
- **Naive:** interesting, manageable design questions

Social Networks

- What are the plausible models?

Thank you!