Network diffusion 2.0?

Anna Scaglione
Bellair workshop 2012

Network Gossiping 1.0

- Two main research threads in networking and systems
  - Broadcasting as “rumor-mongering”
    - Born out of epidemic models, studied in math in the mid 80s gained prominence in engineering (mostly networking) as a method to share content in peer 2 peer networks
    - Peaked in a long period ~ 2000-2007
  - Average Consensus Gossiping
    - Computing averages (or anything that could be written as an average), introduced by Tsitsiklis, became popular in networking, control and signal processing
    - Peaked around 2003-2007

Characteristics of the problem examined

- The tasks considered have modest complexity
  - Distribution and replication
  - Aggregation of data to respond queries

$$q = \sum_{i=1}^{n} f_i(x_i)?$$

- The data are scattered
  - Gossip is a tool to simplify management
  - Analysis of speed, resilience, fault tolerance
- Typical complexity $\Theta(n \log n)$

Impact?

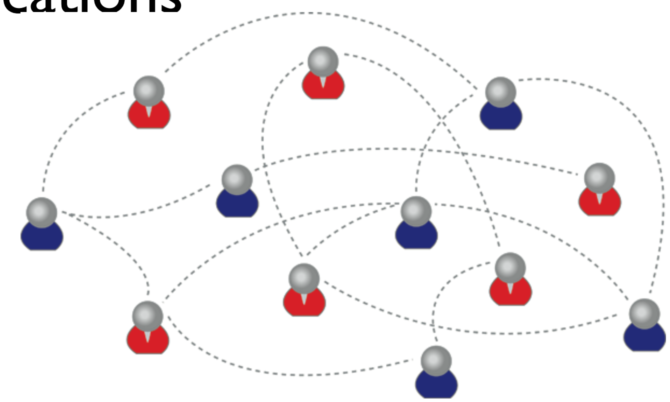
- Video multicasting
  - PPLive a Chinese media company is known to use a form of algebraic gossiping
  - It combines network coding with epidemic replication
- For sensor networks?
- For synchronization?
- In robotics?

Computer Systems Gossiping 1.0

- Consensus is associated with Paxos, by Leslie Lamport
  - a fictional legislative consensus system describes formally a fault tolerant method to attain consensus for distributed processing
  - Aim → consistency in the presence of lousy terminals and links
  - In a typical deployment there is a continuous stream of agreed values acting as commands to a distributed state machine
- Impact?

Network gossiping 2.0

- Option 1 – Design and Analyze applications
  - Learning emerging from interactions between social agents
  - Challenge: Does learning emerge?
  - It is not a green field



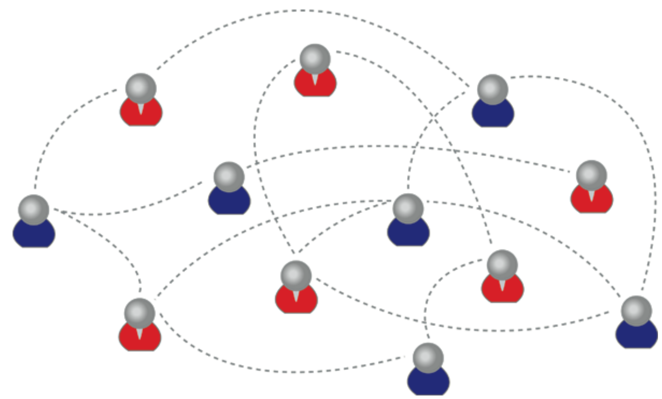
- Option 2 – Solve more complex consensus problems….
  - Optimization in the cloud
  - Challenge: Does learning emerge?
  - Possible application
    - Database transcription of sensor data

Social Learning

Bayesian and non Bayesian analyses

Social learning models

- Objective
  - Understanding how our opinions change when we observe others opinions or actions



- Studies in
  - Social Sciences
  - Physics/Statistical Mechanics
  - Economics
- Mostly analytical
  - Physics/Statistical Mechanics
    - Non linear dynamics of the agent states under plausible interaction models
  - Economics
    - Bayesian learning and self-interested actions

Social Networks in Economics

- Rational Agents
  - Belief = probability of the state of the world θ after observations of private signals s and actions a maximizing expected utility
- Typical cases
  - Learning from binary signals and binary actions

Private information	$s=1$	$s=0$
$\theta=1\vert$	q	$1-q$
$\theta=0\vert$	$1-q'$	q'

Utility $u(a,\theta)=(\theta-c)a, \quad c\in(0,1)$

  - Learning from Gaussian signals

$$\theta\sim\mathcal{N}(m,1/\rho), \quad s\sim\mathcal{N}(\theta,1/\rho_\epsilon)$$

Utility $u(a,\theta)=-(a-\theta)^2$

Private belief

- Given the Gaussian prior $\theta \sim \mathcal{N}(m, 1/\rho), \quad s \sim \mathcal{N}(\theta, 1/\rho_\epsilon)$

Private belief

$$\mathcal{N}(m', 1/\rho')$$

$$1/\rho' = 1/\rho + 1/\rho_\epsilon$$

$$m' = (1-\alpha)m + \alpha s$$

$$\alpha = \rho_\epsilon/\rho'$$

- Given the binary experiment

Private belief

$$x' = \frac{e^{\lambda'}}{1+e^{\lambda'}}$$

$$\lambda' = \lambda + \log\left(\frac{P(s|\theta=1)}{P(s|\theta=0)}\right)$$

The basic analysis

- The action history is public

$$h_t = (a_t, \ldots, a_1)$$

- Either the log-likelihood function or the mean of the density are shifted by all agents by the same amount

$$\lambda_{t+1} = \lambda_t + \nu_t \qquad \nu_t = \log\left(\frac{P(a_t|\theta = 1)}{P(a_t|\theta = 0)}\right)$$

- For the Gaussian case

$$m_t = E[\theta|h_t], \quad 1/\rho_t = MMSE_t(\theta|h_t)$$

One advantage of the Bayesian model

- The model has one distinct advantage in general (no matter what the utility is)
- The sequence of beliefs has to be a martingale:
  - *Updating is rational and rationally anticipated → The expected revision of the belief must be zero*

$$x_t = \mathbb{E}[\theta | h_t] = E[x_{t+1} | h_t]$$

- The Martingale Convergence Theorem (MCT) ensures that the belief must converge to a random value μ^*
- Note it cannot go to 0 or 1 because its next move would be anticipated

Social learning in the Gaussian case

- This case for the classic setup is greatly simplified by the choice of utility, which leads to a Bayesian MMSE estimation problem
- The action that minimize the squared error on the average current belief is

$$a_t \equiv m_{t+1} = E[\theta|h_t] = (1 - \alpha_t)m_t + \alpha_t s_t$$

$$\rho_{t+1} = \rho_t + \rho_\epsilon$$

- The action *reveals the private information* of the agent and therefore the network attains asymptotic the right belief and learning continues indefinitely

The Binary case → the BHW model

- Credited to Bikhchandani, Hirshleifer and Welch (1992)
- Showed *informational cascades* in Bayesian learning
- Recall that the agents maximize the expected utility

$$u(a,\theta) = (\theta - c)a, \quad c \in (0,1)$$

- Hence the optimum action is

$$\max_{a \in \{0,1\}} \left(\mathbb{E}[\theta | h_t] - c\right) a \quad \rightarrow a = u(x_t - c)$$

- Equivalently:

$$a_t = u(\lambda_t - \gamma) \quad \gamma = \log \frac{c}{1-c}$$

Herding

- The probability of a=1 is the CDF of the belief at γ
- Hence the belief evolution is a Markov Chain

$$x_{t+1} = \mathcal{B}(x_t, a_t)$$

$$P(a_t = 1) = 1 - F_t^{\theta}(\gamma)$$

- Proposition (BHW '98)

1. $x^* < x_t < x^{**}$ Agents invests in and only if $s = 1$
2. $x_t > x^{**}$ Agent t invests no matter what s
3. $x_t \leq x^{**}$ Agent t does not invest

- Cascade (2 or 3) will occur almost surely in finite time

Bayesian model challenges

- Except for the Gaussian case with quadratic utility rational herding analyses often requires great mathematical sophistication
  - Proofs are indirect
  - Brute force calculation of the dynamics are difficult
  - Stochastic dominance and MCT are key ingredients
- Some economic papers resort to numerical simulations using neural networks to emulate the learning step
- Experiments have shown that humans are not rational
- Reference:
  - *"Rational Herds: Economic Models of Social Learning" by C. Chamely, Cambridge 2004*

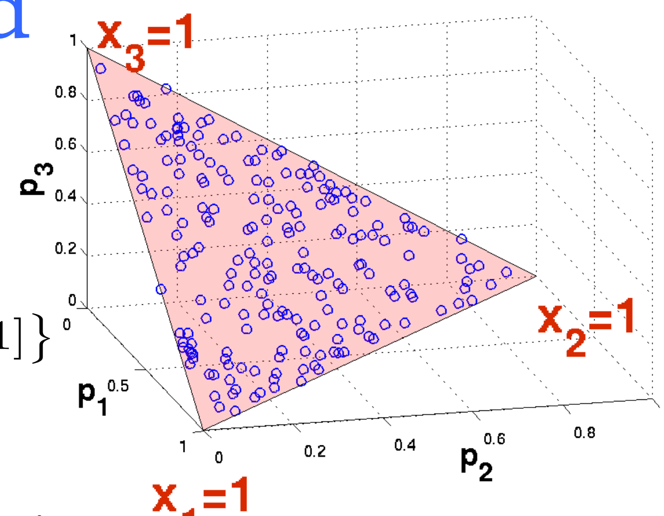
Non Bayesian models

- Statistical physics approach to social dynamics
- Topics studied
  - Dynamics of opinions
  - Cultural dissemination
  - Crowd dynamics
  - Emergence of hierarchies
  - …
- Let's focus on opinion dynamics
  - A rather comprehensive and intimidating tutorial
  - *"Statistical physics of social dynamics"* Claudio Castellano et. al.
    - http://arxiv.org/abs/0710.3256v2

Opinion dynamics models

- Discrete voter model
  - Clifford and Sudbury, 1973 - Holley and Liggett, 1975
    - The agents have a discrete state $x_i \in \{0, 1\}$
    - At random times they copy each other
  - Ercan Yildiz discussed this extensively last year while analyzing the impact of stubborn agents
- Continuous opinion
  - Chatterjee and Seneta, 1977; Cohen et al., 1986; Stone, 1961
    - The agents have a continuous random belief $x_i \in [0, 1]$
  - Bounded confidence model
    - Deffuant et al., 2000 and Hegselmann and Krause, 2002
    - Only if agents have sufficiently similar opinions they mix them

Extensions we have analyzed



- Belief/opinion Model:
  - There are q states of nature
  - Belief: $\mathbf{x} = [x_1, \cdots, x_q]$ (probabilities)
  - Sample space: $\mathcal{X} = \{\mathbf{x} | \sum_i x_i = 1 \text{ and } x_i \in [0,1]\}$
- Opinion Dynamics (non Bayesian):
  - ***After each interaction, the opinion distance cannot increase.***

(a1) $$d_{ij}[k+1] = [1 - \epsilon_k \rho(d_{ij}[k])] d_{ij}[k]$$

stepsize

deg. of opinion change

(a2) $\epsilon_k : \sum_k \epsilon_k = \infty, \ \sum_k \epsilon_k^2 < \infty$

(a3) $\rho(d)$: a non-decreasing function of d.

- Distance measure:
  - $d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j)$: a proper geometric distance
  - $\mathcal{X}$ is bounded w.r.t. the norm $\|\mathbf{x}_i\| := d(\mathbf{x}_i, 0)$
- ***Herding***: $\forall i, j \in V, \ d(\mathbf{x}_i, \mathbf{x}_j) = 0$ ***Polarization:*** non-interacting sub-groups

Interaction Models

$$d_{ij}[k+1] = [1 - \epsilon_k \rho(d_{ij}[k])] d_{ij}[k]$$

- Two classes of interaction models
  - *Soft-Interaction model (=Unequal confidence)*
    *Agents **always** communicate and exchange beliefs*
    - (a4) $\rho(d_{\max}) = \rho_{\min} > 0$
    - (a5) $\rho(d)$ is C^2-differentiable for all d
    - (a6) $\rho(d)d$ is concave
  - *Hard-Interaction model (= Bounded Confidence)*
    No interaction occurs when the distance is greater than or equal to τ (threshold)
    - (a7) $\tau : d \geq \tau \rightarrow \rho(d) = 0$
    - (a8) $\rho(d)$ is C^2-differentiable for $\forall d \in (0, \tau)$
    - (a9) $\rho(0)/\rho(\tau^-) \leq \beta < \infty$
    - (a10) $\rho(d)d$ is concave for $\forall d \in [0, \tau]$

Soft-Interaction Model

- Approach:
  - Step 1: *Stochastic Approximation*

$$d_{ij}[k+1] = [1 - \epsilon_k \rho(d_{ij}[k])] d_{ij}[k]$$

ϵ_k | ϵ_k

$$\dot{d}_{ij}(t) = -\rho(d_{ij}(t)) d_{ij}(t)$$

$$\bar{d} = \sum_{(i,j)\in E} \overline{P}_{ij} d_{ij}$$

$$\dot{\bar{d}} = -\sum_{(i,j)\in E} \rho(d_{ij}) d_{ij}$$

(in expectation)

$$\dot{d}^2_{ij}(t) = -[2\rho(d_{ij}(t)) - \epsilon_k \rho^2(d_{ij}(t))] d^2_{ij}(t)$$

$$\bar{d^2} = \sum_{(i,j)\in E} \overline{P}_{ij} d^2_{ij}$$

$$\dot{\overline{d^2}} = -\sum_{(i,j)\in E} [2\rho(d_{ij}) - \epsilon_k \rho^2(d_{ij})] d^2_{ij}$$

(in mean square)

  - Step 2: *Find the upper bound and the lower bound of the system*
  - Step 3: *Convergence analysis*

- Lemma 1 [Convergence in **expectation**]: *under (a1)-(a6),*
  - $\exists \alpha \in (0, \frac{1}{2}]$ *s.t. the dynamics of* $\bar{d}$ *is upper and lower bounded by* $-\rho(\bar{d})\bar{d} \leq \dot{\bar{d}} \leq -\alpha\rho(\bar{d})\bar{d}$
  - *Local rate of convergence: exponential (dot)* $\alpha[\rho(\bar{d}) - d\rho'(\bar{d})] \leq r(\bar{d}) \leq \rho(\bar{d})\bar{d}\rho'(\bar{d})$

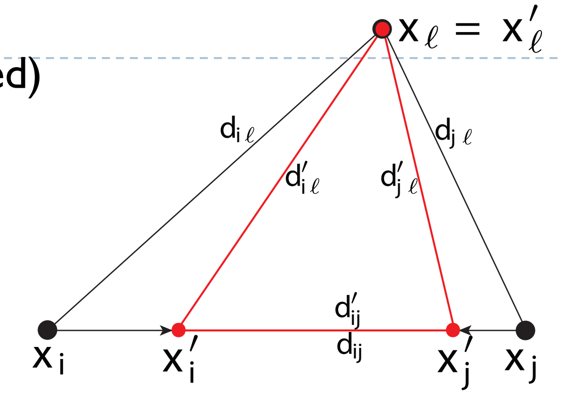
- Lemma 2 [Convergence in **mean square**]: *under (a1)-(a6),*
  - $\exists \tilde{\alpha} \in (0, \frac{1}{2}]$ *s.t. the dynamics of* $\overline{d^2}$ *is upper and lower bounded by* $-2\rho(\sqrt{\overline{d^2}})\overline{d^2} \leq \dot{\overline{d^2}} \leq \rho(\sqrt{\overline{d^2}})\overline{d^2}$
  - *Local rate of convergence: exponential*

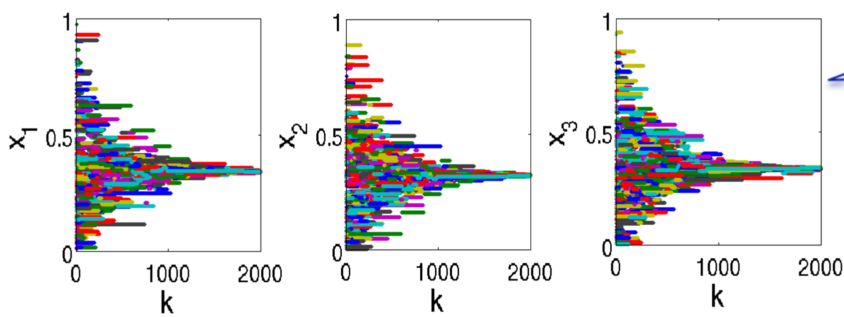
Simulations

- $G_c = G(n, \alpha)$ with $n = 100,\ \alpha = 0.6$ (connected)
- L2 distance metric
- deg. of opinion change: $\rho(d) = 0.5 - 0.2d$
- Beliefs are updated through the shortest path; if $h_{ij} \neq 1$

$d(\mathbf{x}_i, \mathbf{x}'_i) = \mu(d_{ij}) d_{ij}$ and $d(\mathbf{x}_j, \mathbf{x}'_j) = \gamma(d_{ij}) d_{ij}$

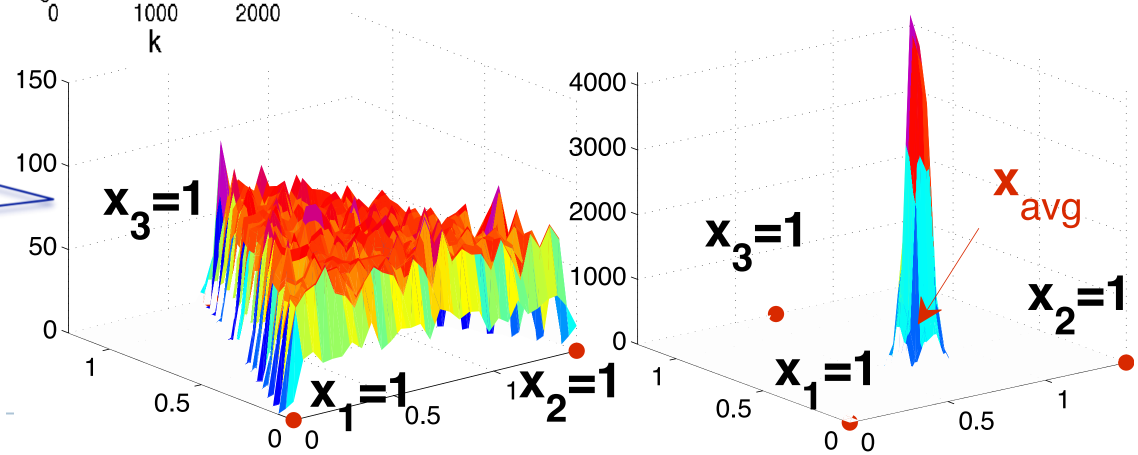
$$\rho(d_{ij}) = \mu(d_{ij}) + \gamma(d_{ij})$$





Evolution of Opinion Distribution: *Each line segment corresponds to a node and a segment terminates when a node interacts and changes its belief.*

A **histogram** of 300 belief profiles at time zero (left) and after the dynamics have stabilized (right.)



Hard-Interaction Model (Deffuant)

- Step 1: Stochastic Approximation

$$d_{ij}[k+1] = [1-\epsilon_k \rho(d_{ij}[k])]d_{ij}[k]$$

ϵ_k

$$\dot{d}_{ij}(t) = -\rho(d_{ij}(t))d_{ij}(t)$$

$\bar{d} = \sum_{(i,j)\in E} \overline{P}_{ij} d_{ij}$

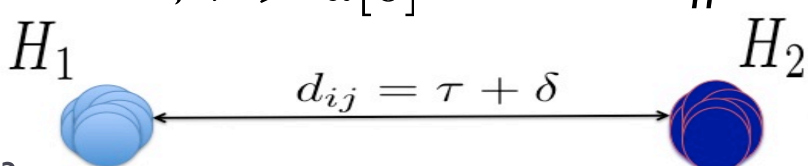
$$\dot{\bar{d}} = -\sum_{(i,j)\in E} \rho(d_{ij})d_{ij}$$

  - Step 2: Find the lower bound system $\longrightarrow$ $\dot{\bar{d}} \geq -\beta\rho(\bar{d})\bar{d}$
  - Step 3: Analyze the lower bound system $\dot{b} = -\beta\rho(b)b$

  Lemma 3: *Under (a1)-(a3) and (a7)-(a10)*

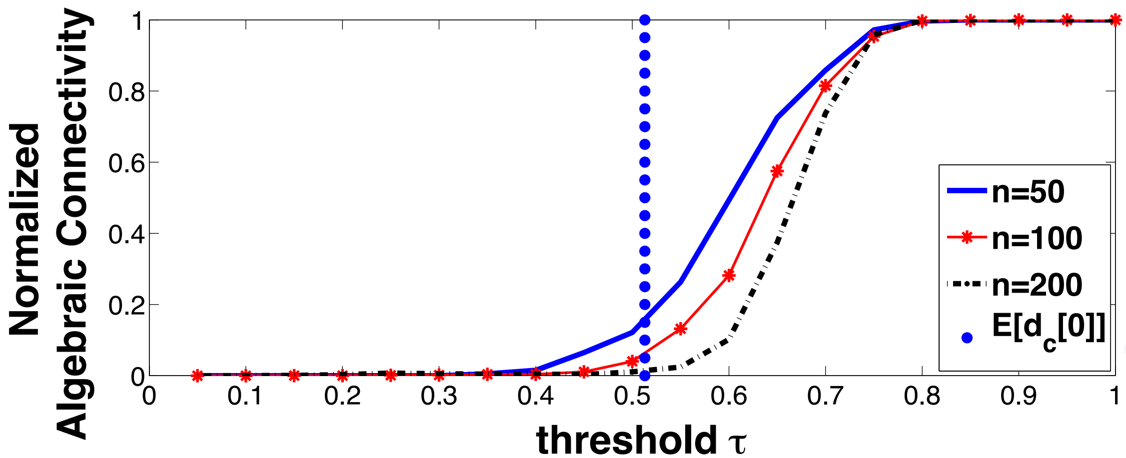
  the system $\dot{b} = -\beta\rho(b)b$ *converges if* $\tau > b(0)$.

- Lemma 4: *A **necessary** condition for the system to converge almost surely is* $\tau > d[0]$.
- However, $\tau > d[0]$ *is **not** the sufficient condition.*



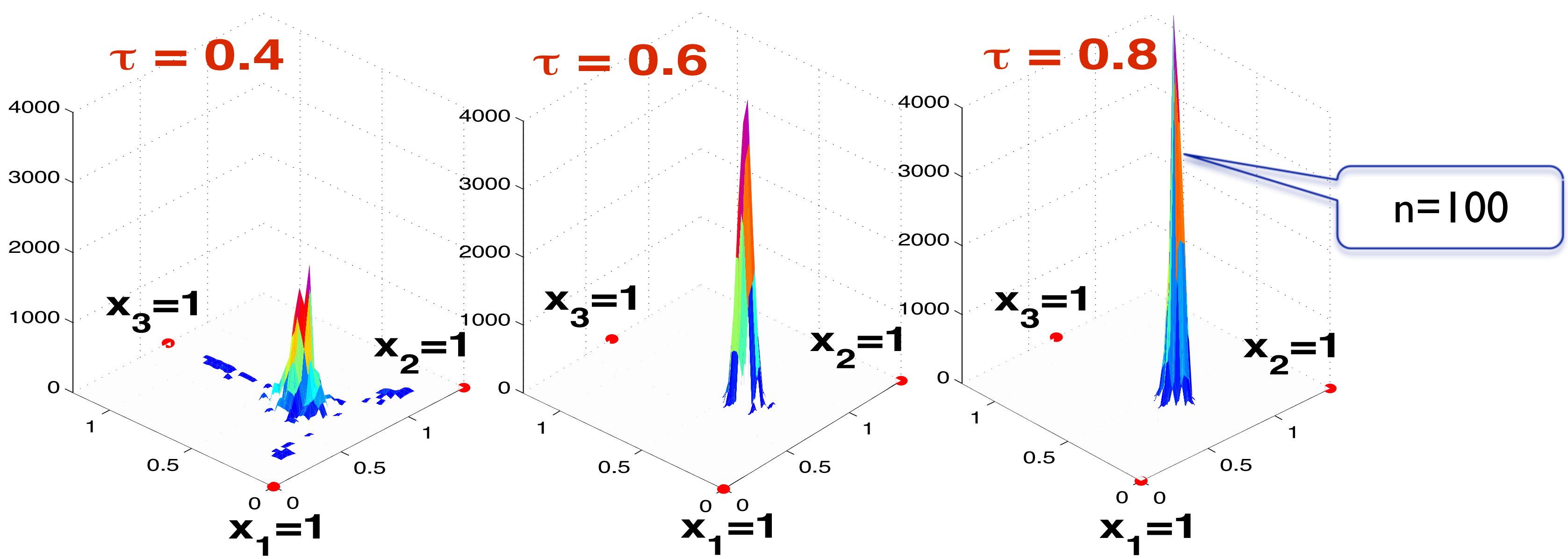
If $\sum_{(i,j)\in H_1\times H_2} \overline{P}_{ij} < \frac{\tau}{\tau+\delta} \Rightarrow \bar{d}[0] < \tau$

Simulations (Necessary Condition)



Averaged over 400 trials. Each starts with an uniform random initial belief profile and

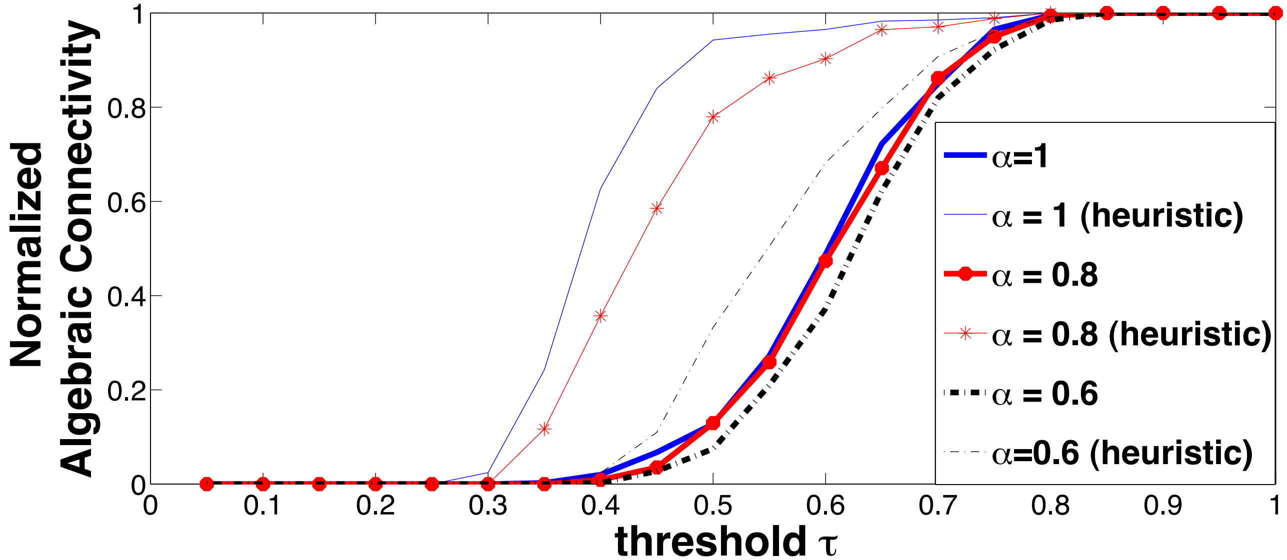
$$\rho(d) = 1 \ \forall d < \tau.$$



Is the Rate of Interaction Important?

- We also studied how the social fabric affects the herding behavior.
  - Is there any particular distribution for $\overline{P}_{ij}$ that favors agreement vs. disagreement?
  - The following simple lemma establishes the necessary condition proved in Lemma 4 is not affected by the social fabric as long as $\overline{P}_{ij}$ are drawn independently from the beliefs.
- **Lemma 5:** *If the social fabric (represented by $\overline{P}_{ij}$) is **random** and **independent** of $d_{ij}[0]$ then it will, on average, exhibits the same phase transition.*
- **Local rewiring topology:**
  - Aim: to decrease $d[0]$
  - How? --- by choosing an opinion-dependent $\overline{P}_{ij}$
  - Topology: (1) remove links between agents whose $d_{ij}[0] > \tau$
    (2) redistribute $\overline{P}_{ij}$ uniformly to the remaining neighbors

Simulations (Local Rewiring Topology)



- *Observations:*
  - Performances are similar when the underlying network is independent of $d_{ij}[0]$.
  - When $\overline{P}_{ij}$ is correlated with $d_{ij}[0]$ through the local rewiring topology, the probability of forming a convergent belief increases.

Considerations

- Both point of views are exciting to learn and fun to analyze when they do not give me an headache
- My worry is:
  - What is our value added?
- I tend to agree with considering non-Bayesian models not sufficiently based on a "rational" argument → tenuous connection with evidence
- Non Bayesian models also do not capture the selfishness that triggers the action and therefore the information exchange
- Attacking the Bayesian models tome has more profound implications, also in terms of automation of decisions
  - Much harder!

Optimization via Network diffusion

Where and when….

Optimization via network diffusion

- Clearly an engineering problem - typical sensing problem

$$z_i = f_i(x) + v_i$$

- Non-linear least square

$$\hat{x} = \text{argmin}_x \sum_{i=1}^{n} J_i(x) = \sum_{i=1}^{n} \|z_i - f_i(x)\|_{R_i}^2$$

- Idea (Angelia is the one that can elaborate) approximate the gradient descent for the global objective

$$J_0(x) = \sum_{i=1}^{n} J_i(x) \approx const. + \sum_{i=1}^{n} \|x - x^*\|_{H_i}$$

Local surrogate of the global function

- Make a surrogate with two terms

$$J_0(x) \approx_i \sum_{j \in \mathcal{N}_i} J_i(x) + \sum_{j \in \mathcal{N}_i} \|x - \hat{x}_j^*\|_{H_i}$$

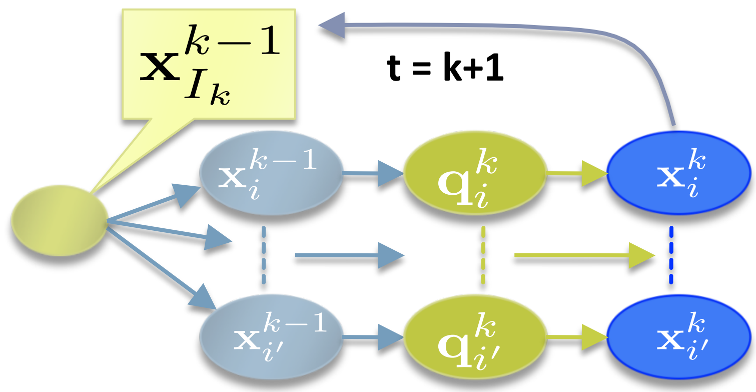
- Consensus step to decrease $\sum_{j \in \mathcal{N}_i} \|x - \hat{x}_j^*\|_{H_i}$
- Gradient descent step to decrease $\sum_{j \in \mathcal{N}_i} J_i(x)$
  - Angelia has the strongest results for convex functions

Technical Detail

- **Decentralized Formulation :**

$$\min_{\mathbf{x}} J(\mathbf{x}) = \sum_{i=1}^{I} J_i(\mathbf{x}) = \sum_{i=1}^{I} [\mathbf{z}_i - \boldsymbol{f}_i(\mathbf{x})]^T \mathbf{R}_i^{-1} [\mathbf{z}_i - \boldsymbol{f}_i(\mathbf{x})]$$

- **Iterative Updates by Asynchronous Gossiping:**



Broadcasted from Agent I_k at time k

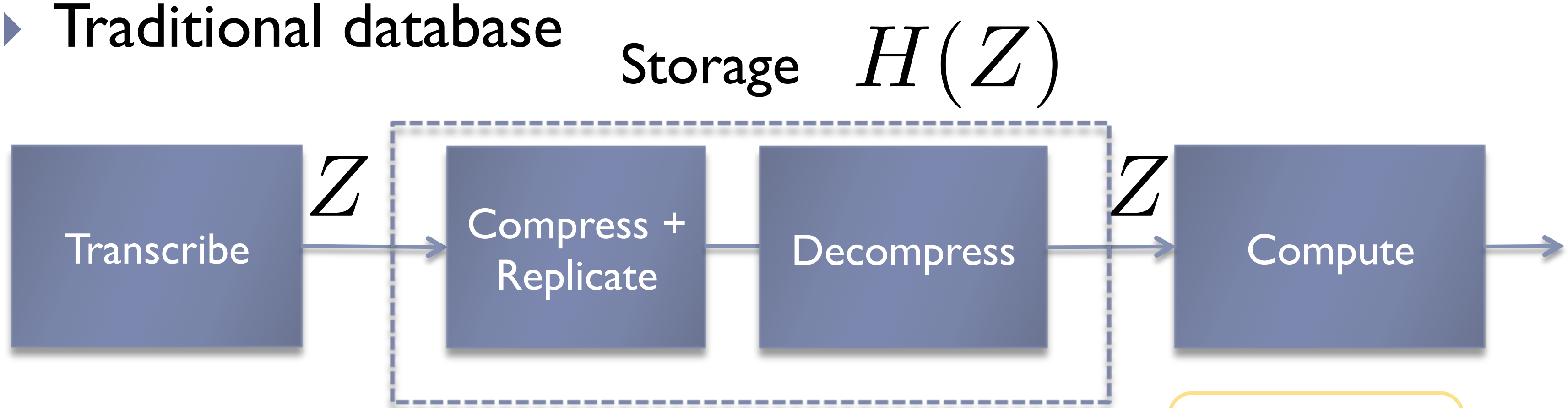
$$\begin{aligned} \mathbf{q}_i^k &= \beta \mathbf{x}_{I_k}^{k-1} + (1-\beta)\mathbf{x}_i^{k-1} \\ \mathbf{x}_i^k &= \mathbf{q}_i^k - \alpha_{i,k}\mathbf{d}_{i,k} \end{aligned} \qquad i \in \mathcal{N}(I_k)$$

Update = function descent of local cost $J_i(\mathbf{x})$

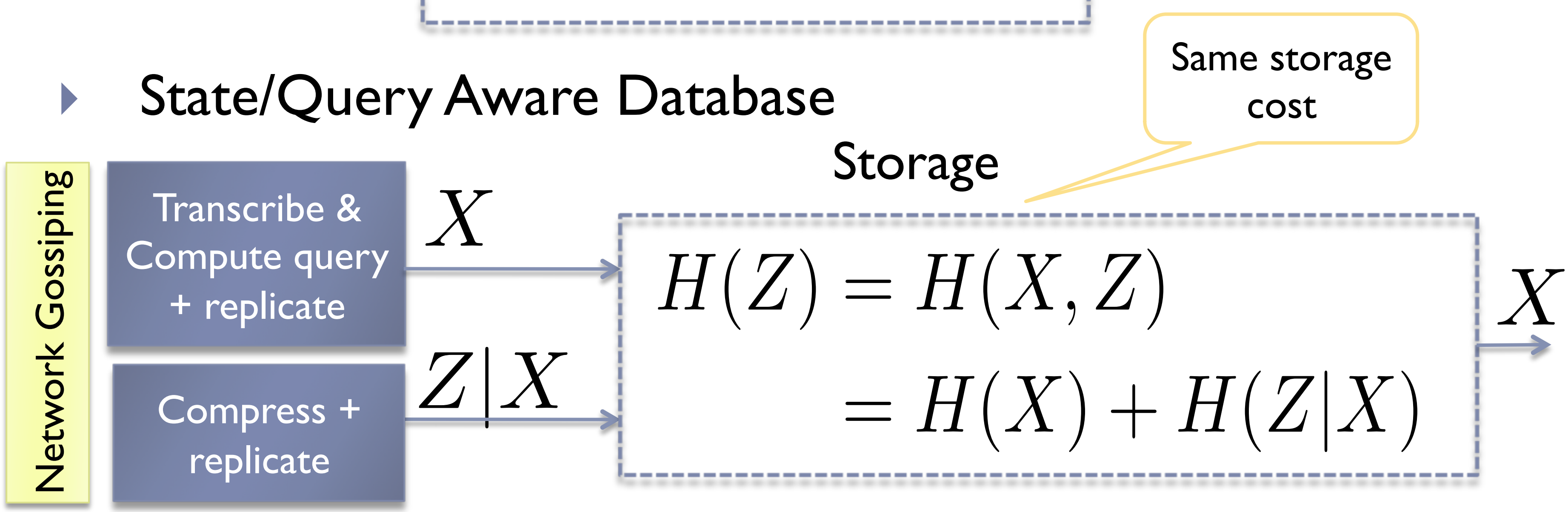
- **Application: Cyber Physical Systems**
  - Today the model is Sensory Control and Data Acquisition (SCADA)
  - State estimation is done after transcription in the database system…

Network gossiping as a transcription tool

- Traditional database



- State/Query Aware Database



Measurement Compression

- Quantized measurement data

snapshot index

$$\boldsymbol{q}_{i,m}[\ell] = \mathcal{Q}\left(z_{i,m}[\ell]\right), \quad \mathcal{Q}(\cdot) : \mathbb{R} \rightarrow \{0,1\}^L$$

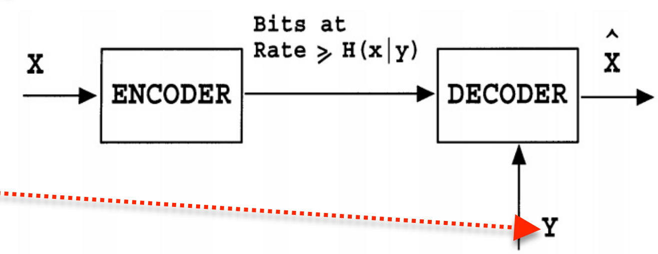
- Accurate state implies accurate pseudo-measurements

$$\|\widehat{\mathbf{x}}_j[\ell] - \widehat{\mathbf{x}}_i[\ell]\| \rightarrow 0 \longrightarrow \|f_{j,m}\left(\widehat{\mathbf{x}}_j[\ell]\right) - f_{j,m}\left(\widehat{\mathbf{x}}_i[\ell]\right)\| \rightarrow 0$$

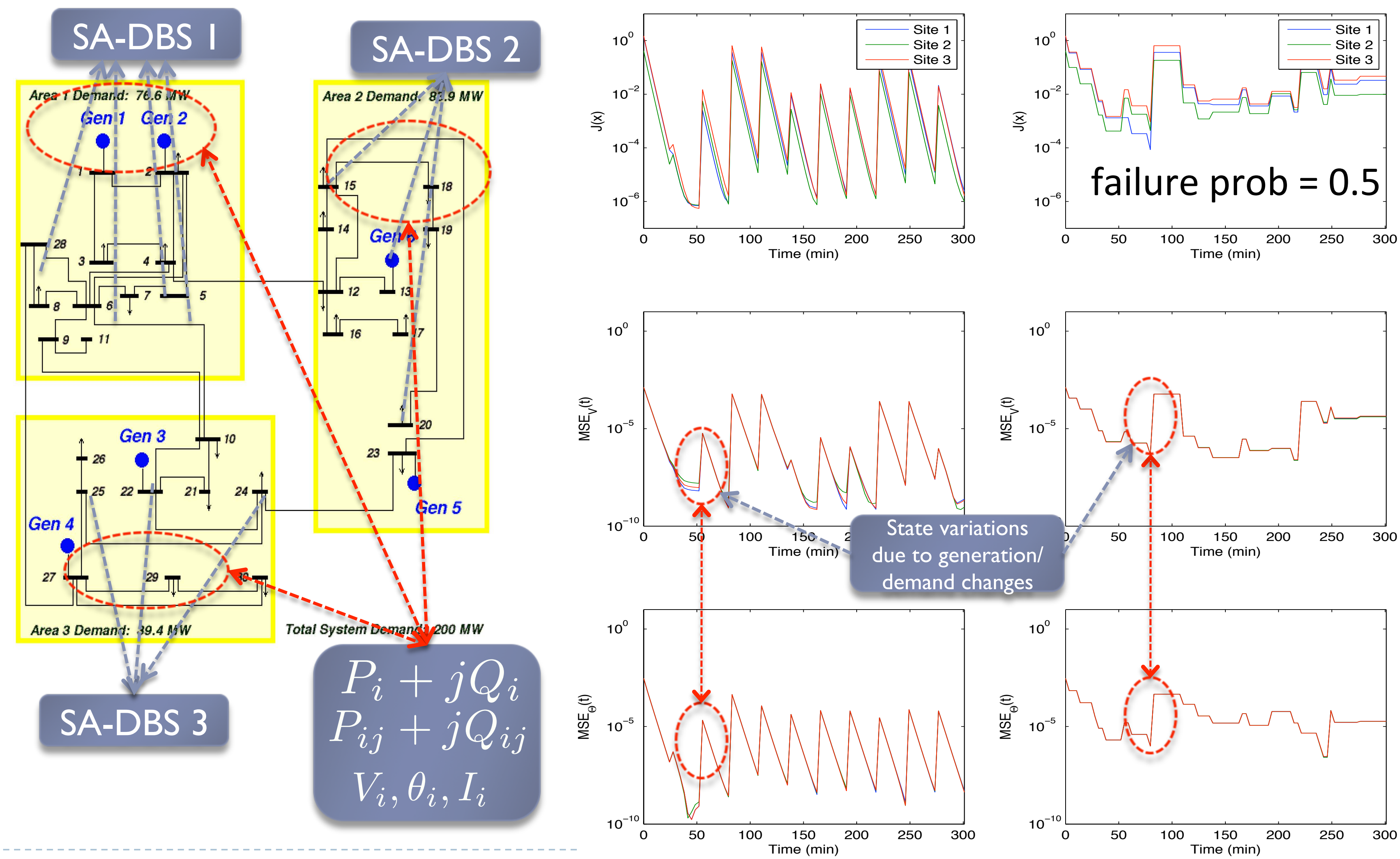
- Pseudo-measurements serve as side information

$$\begin{aligned} z_{j,m}[\ell] &= f_{j,m}\left(\mathbf{x}^\star[\ell]\right) + \varepsilon_{j,m}[\ell] \\ &\stackrel{(\star)}{=} f_{j,m}\left(\widehat{\mathbf{x}}_i[\ell]\right) + \mathcal{O}\left(\|\widehat{\mathbf{x}}_i[\ell] - \mathbf{x}^\star[\ell]\|\right) + \varepsilon_{j,m}[\ell] \\ &\approx \zeta_{j,m}^{(i)}[\ell] \end{aligned}$$

Few Errors $\boldsymbol{q}_{j,m}^{(i)}[\ell] = \mathcal{Q}\left[\zeta_{j,m}^{(i)}[\ell]\right]$

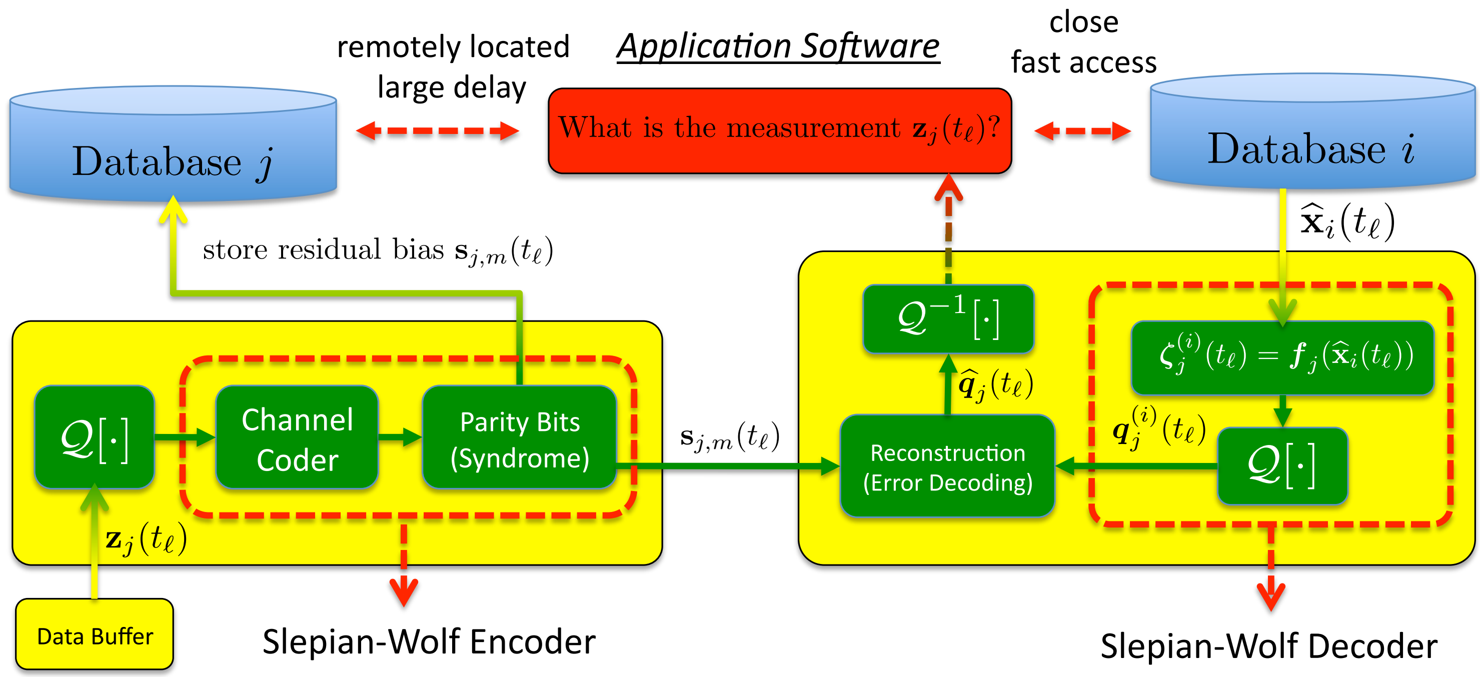


SE via gossiping on IEEE 30 Bus System



Example of Measurement Compression

- **State-Aware Slepian-Wolf Codes (SA-SWC)**

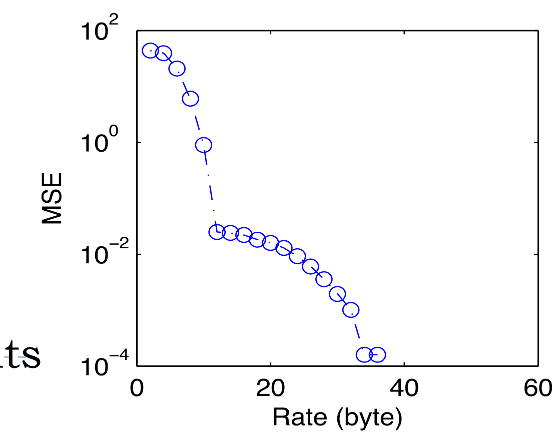


Example:

- IEEE 30 bus system, approximately 250 quantized measurements with L=8 bits

- Use (n, n-2t) Reed Solomon codes, t= 1:60 with 8-bit symbols (one byte per meas.)

- 30 bytes suffice at each DDBS to recover the measurements (30*3/250 = 36%)



Considerations

- Societies filter information via message passing, hence they are naturally interesting for those who research network gossiping
- Unfortunately social learning is not a green field
  - Rational agents models are far more interesting but far more complex
- Advanced network gossiping techniques for sensor data are possibly going to have more impact if they are integrated with the transcription of data into an archive
- They are powerful methods to compute and disseminate answers to queries, which could populate the database first