A message-passing approach to stochastic optimization and inverse dynamical problems

Fabrizio Altarelli
Alfredo Braunstein
Luca Dall’Asta
Riccardo Zecchina

@Politecnico di Torino, Italy
optimization under uncertainty

(i)

\[ x_1^* = \arg \min_{x_1} \mathbb{E}_t \min_{x_2} \mathcal{E}(x_1, x_2, t) \]
\[ x_i^* = \arg \min_{x_i} \mathbb{E}_{t_i} \min_{x_{i+1}} \mathbb{E}_{t_{i+1}} \cdots \min_{x_k} \mathbb{E}_{t_k} \mathcal{E}(x, t) \]

(ii)

\[ X^{t=0} \rightarrow f(X^t) \rightarrow X^{t=T} \]

irreversible process

? e.g. induce large deviations in Bootstrap percolation
viral marketing & social networks

signaling pathways

epidemics, fire extinction models

complex dynamics (glassy dynamics), neural systems, ...
plan of the talk

• Brief review: sampling & optimization via BP and SP equations

• Stochastic Optimization & Multi-level optimization

• Inverse dynamical problems & control of network cascade processes
we start with a simple to state problem (but hard to solve)

$$\text{GF}(2)$$

$$\begin{cases} 
A \sigma = 0 \\
\min d(\sigma, \tau)
\end{cases}$$

$$A : m \times n \text{ matrix}, \quad \alpha = \frac{m}{n}$$

$$d(\sigma, \tau) = \sum_{i}(1 - \delta(\sigma_i, \tau_i))$$
A is sparse and random, \( \tau = (0, 0, \ldots, 0, 0) \)
Upper bound by first moment

\[ P(Z \geq 1) \leq E(Z) \quad \text{Markov's inequality (Z non-negative)} \]

Bound for the number of pair of solutions at distance x:

\[ Z(x) = \sum_{\bar{\sigma}, \bar{\tau}} \delta \left( d_{\bar{\sigma}, \bar{\tau}} \in [Nx + \epsilon(N), Nx - \epsilon(N)] \right) \delta [\bar{\sigma}, \bar{\tau} \in S] \]

\[ \lim_{N \to \infty} \epsilon(N)/N = 0 \]

\( S \) is the set of solutions
$\frac{1}{N} \log \bar{N}$

?
Lower bound by weighted second moment

\[ P(Z > 0) \geq \frac{E(Z)^2}{E(Z^2)} \quad (Z \geq 0) \quad \text{Chebyshev inequality} \]

\[ Z = \sum_{\vec{\sigma}, \vec{\tau}} \delta(d_{\vec{\sigma}, \vec{\tau}} = xN)W(\vec{\sigma}, \vec{\tau}) \]

\[ Z > 0 \leftrightarrow \text{existence of pairs of solutions} \]
Statistical physics of random CSP in brief

1. $\alpha_d < \alpha < \alpha_c$  Gibbs measure decomposition in clusters, various forms of hardness for algorithms

2. $\alpha < \alpha_d$  solution space is well connected, algorithms are effective ($t=\text{P}(n)$)

complete proofs for random XORSAT (pspin)
structure of solution-space
pure states decomposition of $\mu(\vec{\sigma})$.

\[ w_\gamma = \sum_{\vec{\sigma} \in A_\gamma} \mu(\vec{\sigma}) \]

$w_1 > w_2 > w_3 > \ldots$

- **RS**: most of the measure in a single cluster
\[
\lim_{N \to \infty} w_1 = 1
\]

- **d1RSB**: the measure divides in $e^{N\Sigma^*}$ clusters

- **1RSB**: the measure condensates in sub-exp number of clusters
\[
\lim_{n \to \infty} \lim_{N \to \infty} \sum_{i=1}^{n} w_i = 1
\]
Cavity method & Loopy Belief Propagation equations

\[
\{ P_{j \rightarrow b}(\sigma_i) \}
\]

\[
P_{i \rightarrow a}(\sigma_i)
\]

total graph
Exact if correlation decay holds

\[
P(\sigma_1, ..., \sigma_n) = \prod_a P_a(\sigma_i, \sigma_j, \sigma_k) \prod_i P_i(\sigma_i)^{1-n_i}
\]

Message-passing (MP) equations

\[
p_{i\rightarrow a}^{\sigma} \propto \prod_{b \neq a} \eta_{b\rightarrow i}^{\sigma}
\]

\[
\eta_{a\rightarrow i}^{\sigma} \propto \sum \prod p_{j\rightarrow a}^{\sigma_j} \chi[\{\sigma_j\}, \sigma_i]
\]
Multiple clusters of optimal configurations

Main assumptions:

1) # Clusters: \( \mathcal{N}(e) \sim \exp(N\Sigma(e)) \)

2) Correlation decay within clusters: \( \{P_i^{(\alpha)}(\sigma_i)\}_i \)

Eqs. for marginals over BP solutions (Survey Propagation)

\[
\{\{P_i^{(\alpha)}(\sigma_i)\}_i\}_\alpha \rightarrow \mathcal{P}[P]
\]
Sparse graphs: random K-Satisfiability, Graph Coloring, Reconstruction, Graph Theory, Quantum Spin systems, Source Coding, matching, compressed sensing ... rapidly growing list

Dense graphs: Inverse Problems (e.g. Ising/Potts), Machine Learning, Clustering, ...

Graphs with topological constraints (Steiner trees, optimal subgraphs, ...)

Stochastic Optimization ...

I should quote a lot of people from CS, math and physics!
some problems left open

- building an algorithmic theory of statistical physics algorithms: e.g. upper and lower bounds for loopy MP
- coupling MP with other classes of algorithms
- hard random problems: the barrier of frozen variables (e.g. packing problems)

the general mathematical/algorithmic framework is highly incomplete
MP methods can be extremely effective for approximate SAMPLING and COUNTING

- e.g. RSB: MP > MC

- extremely fast approximation techniques

Cost-functions defined through sampling/counting
Let’s now consider an ensemble of CSP with random components assuming that correlation decay holds for almost all the realizations.

Eqs. for marginals over BP solutions over the random ensemble

\[
\left\{ \left\{ P_i^{(\alpha)}(\sigma_i) \right\}_i \right\}_\alpha \rightarrow \mathcal{P}[P]
\]

random realizations
Can be used to compute statistics over minima of a given random function

i) Functional cavity approach, Survey Propagation (SP, Pop. Dyn.) eqs

ii) sampling (e.g. with MCMC)

For complex landscapes SP can be more efficient than sampling

For k-stage optimization, sampling can just be unfeasible

In same cases there is a pseudo-analytic solution: Stochastic Matching
A cavity / mean-field approach to:

• **Stochastic optimization**
  - example: the case of matching

• **Multi-level optimization**
  - example: *inverse dynamics*
Stochastic optimization: General Problem

Given a cost function $\mathcal{E}(\mathbf{x}, \mathbf{t})$ that depends on $n$ variables $\mathbf{x} \equiv \{x_1, \ldots, x_n\}$ and $m$ independent random parameters $\mathbf{t} \equiv \{t_1, \ldots, t_m\}$ with distr. $\mathcal{P}_i(t_i)$

- **Step (i):** the optimizer has to decide the value of variable $x_i$ and afterwards variable $t_i$ is picked from $\mathcal{P}_i(t_i)$.

- **Step (n):** At the end of the process, all $\mathbf{x}$ and $\mathbf{t}$ result fixed and the value of $\mathcal{E}$ is computed.

**GOAL:** minimize the average final value of $\mathcal{E}$

*“$x_i$ creates an infrastructure for $t_i$“*
More precisely:

Variables and parameters are separated in $k$ stages:

$$x = x_1 \cup x_2 \cup \cdots \cup x_k, \quad t = t_1 \cup t_2 \cup \cdots \cup t_k$$

Minimum expected final energy:

$$x^*_i = \arg \min_{x_i} \mathbb{E}_{t_i} \min_{x_{i+1}} \mathbb{E}_{t_{i+1}} \cdots \min_{x_k} \mathbb{E}_{t_k} \mathcal{E}(x, t)$$

Minimum expected maximal energy (over $t$):

$$\hat{x}_i = \arg \min_{x_i} \max_{t_i} \min_{x_{i+1}} \max_{t_{i+1}} \cdots \min_{x_k} \max_{t_k} \mathcal{E}(x, t)$$

Maximum probability of property $P(x,t)$:

$$\tilde{x}_i = \arg \max_{x_i} \mathbb{E}_{t_i} \max_{x_{i+1}} \mathbb{E}_{t_{i+1}} \cdots \max_{x_k} \mathbb{E}_{t_k} 1[\mathcal{P}(x, t)]$$
Notable cases:

\[ k = 1 \quad \rightarrow \quad \text{Optimization of a stochastic cost function} \]

\[ k = 2 \quad \rightarrow \quad \textbf{Two-stage problem: } x_1 \text{ creates the infrastructure, stochastic events } t, \ x_2 \text{ optimizes the final outcome} \]

\[ x_1^* = \arg\min_{x_1} \mathbb{E}_t \min_{x_2} \mathcal{E}(x_1, x_2, t) \]

(can be NP-hard even if \( k=1 \) is polynomial)

\[ k = n \quad \rightarrow \quad \textbf{on-line multi-stage problem} \ (“game against nature”), alternating process \]

\[ x_i^* = \arg\min_{x_i} \mathbb{E}_{t_i} \min_{x_{i+1}} \mathbb{E}_{t_{i+1}} \cdots \min_{x_k} \mathbb{E}_{t_k} \mathcal{E}(x, t) \]

(PSPACE, approximations needed)
Introduce replicas: Sample Average Approximation (SAA)

\[ \mathcal{E}(x_1, x_2; t_2) \rightarrow \bigcup_{\alpha=1}^{M} \mathcal{E}(x_1, x_2^\alpha; t_2^\alpha) \]

\[ X = \{x_1, x_2^1, x_2^2, \ldots, x_2^\alpha, \ldots, x_2^M\} \]

Computational cost: \( O(f(N)) \rightarrow O(f(NM)) \)

If the energy landscape is complex, \( M \) has to be large to provide accurate results

Not efficient for large scale problems
our approach for two-stage problems

1. \[ \arg \min_{x_2} \mathcal{E}(x_1, x_2, t) \rightarrow \mathcal{E}_1(x_1, t, u) \]

   u are marginals of x_2 & satisfy the T=0 equations

2. \[ \mathbb{E}_t \mathcal{E}_1(x_1, t, u) \rightarrow \mathcal{E}_2(x_1, U) \]

   use marginals U of the u induced by randomness, equations for the U are local (SP equations).

3. \[ \arg \min_{x_2} \mathcal{E}(x_1, x_2, t) \rightarrow \mathcal{E}_1(x_1, t, u) \rightarrow \mathbb{E}_t \mathcal{E}_1(x_1, t, u) \rightarrow \mathcal{E}_2(x_1, U) \rightarrow \mathcal{E}_3 \]

   Minimization over x_1, by introducing marginals of x_1 and of U (marginals of marginals !).
First stage

Second stage

FINAL ENERGY = \( \bigcirc + \bigcirc + \bigcirc \)
optimal solution

FINAL ENERGY =
\[ x \equiv \{x_{lr}, \forall (lr) \in E \} \]

\[ x_{lr} \in \{0, 1\} \]

Diagram:

\[ t_l = 1 \]

\[ \sum_r x_{lr} \leq 1 \forall l \]

\[ \sum_l x_{lr} \leq 1 \forall r \]

\[ \mathcal{E}(x, t) = \sum_l \mathbb{1}[(t_l = 1) \land (\sum_r x_{lr} = 0)] + \sum_r \mathbb{1}[\sum_l x_{lr} = 0] \]

Two-stage problem:

\[ x_1^* \equiv \arg \min_{x_1} \mathbb{E}_{t_2} \min_{x_2} \mathcal{E}(x_1, x_2, t_2) \]
Level I: T=0 Cavity equations

\[
\begin{align*}
    u_{lr} &= - \max \left[ -1, \max_{r' \neq r} h_{lr'} \right] \\
    h_{lr} &= - \max \left[ -1, \max_{l' \neq l} u_{l'r} \right].
\end{align*}
\]

\[
\mathcal{E}^*(x_1, t_2) \equiv \min_{x_2} \mathcal{E}(x_1, x_2, t_2) =
\mathcal{E}(x_1) - \sum_{l \in \mathcal{L}_2(t_2)} \max_r h_{lr} - \sum_{l \in \mathcal{L}_2(t_2)} \max_r u_{lr} + 2 \sum_{(lr): l \in \mathcal{L}_2(t_2)} 1[h_{lr} + u_{lr} = 2]
\]

Level II: Survey Propagation equations

\[
\begin{align*}
    U_{lr} &\equiv \mathbb{P}[u_{lr} = 1] \, , \, H_{lr} \equiv \mathbb{P}[h_{lr} = 1] \\
    U_{lr} &= \mathbb{P}[t_l = 1] \mathbb{P} \left[ - \max \left( -1, \max_{r' \neq r} h_{lr} \right) = 1 \mid t_l = 1 \right] = p_l \prod_{r' \neq r} (1 - H_{lr'}) \\
    H_{lr} &= \prod_{l' \neq l} (1 - U_{l'r})
\end{align*}
\]

\[
\mathcal{E}^*(x_1) \equiv \mathbb{E}_{t_2} \mathcal{E}^*(x_1, t_2) = \sum_l p_l \left[ 2 \prod_r (1 - H_{lr}) - 1 \right] + \sum_r \left[ 2 \prod_l (1 - U_{lr}) - 1 \right] + 2 \sum_{(lr)} H_{lr} U_{lr}
\]

T=0 cavity equations (BP,Max-Sum) & Survey-propagation equations
Third level : $P[P[h]]$

Level III: Minimization of $\mathcal{E}^*(x_1)$ (T=0 cavity eqs over messages)

\[ U_{lr}(U) \equiv P[U_{lr} = U] \quad (U_{lr} \equiv P[u_{lr} = 1]) \]
\[ H_{lr}(H) \equiv P[H_{lr} = H] \quad (H_{lr} \equiv P[h_{lr} = 1]) \]

\[ U_{lr}(U_{lr}) = \max_{(10)} \left[ -2(1 - U_{lr}) \prod_{r' \neq r} (1 - H_{lr'}) + 1 - 2U_{lr} + \sum_{r' \neq r} H_{lr'}(H_{lr'}) \right], \quad (\forall l \in L_1) \]
\[ \{H_{lr'} \in \{0, 1\}, r' \neq r\} \text{ such that } U_{lr} + \sum_{r' \neq r} H_{lr'} \leq 1 \]

\[ \cdots \text{(system of coupled functional local equations)} \]

At the fxp:

\[ x_{lr} = 1 \text{ iif } [U_{lr}(1) - U_{lr}(0)] + [H_{lr}(1) - H_{lr}(0)] + 2 > 0 \]
The average minimum energy is then computed by averaging the contributions of all remaining terms. The average of all remaining terms is computed similarly. The contribution of one incoming value of the variable is then obtained using the update equations and where necessary, the marginalization constraints. The values of the incoming messages are obtained as usual for the MS algorithm. In fact, when there is only one incoming value of the variable, the naive procedure would give to each of them a wrong weight. The vertical line is at 0.00, 0.02, 0.04, 0.06, and 0.08. The horizontal line is at 0.00, 0.02, 0.04, 0.06, and 0.12. The Cost is given on the y-axis, and the <degree> is given on the x-axis.
Limitations of LP: \( n=2000, p_i \in [0,1] \)

- SP-Derived
- CPLEX

![Graph showing limitations of LP](image)
Multi-level optimization

1) use MP to evaluate an approximate probability measure over the set of solutions of the underlying problem

2) combined optimization of the final cost function over the variables and the MP measure (messages)

(similar scheme with different MF approximations)

The definition of the optimization problem includes the solution of another underlying optimization problem.
The average price of Anarchy

1) stochasticity --> distribution over pure Nash equilibria

2) optimize the “average” (over pure Nash) social utility

Braess paradox

Boston
Inverse deterministic dynamics for irreversible processes

**Problem**: Given an irreversible dynamical process over a graph, find a minimum set of sites to be activated at time $t=0$ such that a given final state is reached at time $T$.

**Multi-level nature**

*Level 1* the underlying optimization problem is the direct dynamics: Satisfy the local dynamical constraints and minimize the number of active nodes (to avoid self-sustained loops).

*Level 2* among solutions to the direct problem find those of minimum size at $t=0$, with given property at time $t=T$. 
Activating extreme trajectories
(e.g. in Bootstrap Percolation)
$<k> = 5$, $\vartheta = 4$, $N = 80$, $|S| = 31$, $T = 9$
deterministic dynamics in graphs

- Graph of connections $G = (V, E)$, $\partial i = \{ j : (i, j) \in E \}$
- Instantaneous state of the system at time $t$ given by $x_i^t$ for $i \in V$
- Deterministic rule $x_i^{t+1} = F(\{x_i^t\}, \text{params} ...)$
Direct and Inverse problems

Direct problems:
- Given a distribution of initial states $P(x^0)$, compute observables in time, asymptotics, etc.

Inverse problems
- Fix conditions on some $t > 0$ and infer something about $t = 0$, e.g:
  - Characterize the distribution on the initial state $P(x^0)$ given evidence on the final state $P(x^\infty)$
  - Find parameters that drive the system to desired states, possibly fast!

Main idea: describe trajectories by a “static” model
In practice, both are very difficult! Stems from the fact that even distribution of single-site trajectories are exponential objects.
Spread Problems

- **Direct:**
  - Given an initial condition (e.g. seed density), compute observables on the final state (e.g. final density)

- **Inverse:**
  - Find an initial state (e.g. minimal configuration of seeds) that produces some desired final state (e.g. all active)
  - Study large deviations
  - Minimize both the cost of seeds and the cost of inactive sites

\[ E(t) = \sum_i E_i(t_i) = \sum_i \varepsilon_i \delta(t_i, \infty) + \mu_i \delta(t_i, 0) \]

\[ \phi_i(t_i) = e^{-\mu_i \delta(t_i, 0) - \varepsilon_i \delta(t_i, \infty)} \]
An irreversible process

- We will focus on a simple irreversible process called *Bootstrap* problem

\[
x_{i}^{t+1} = \max \left\{ x_{i}^{t}, \mathbb{I} \left[ \sum_{j \in \partial i} w_{ji} x_{j}^{t} \geq \theta_{i} \right] \right\}, \quad t \geq 0
\]

- We assume $\theta_{i}, w_{ij} \in \mathbb{N}_{0}, x_{i}^{t} \in \{0, 1\}$

- Non-decreasing $x_{i} \implies$ compact trajectory parametrization:
  - $x_{i} = (0, \ldots, 0, 1, \ldots) \longrightarrow t_{i} = \min \{ t : x_{i}^{t} = 1 \}$ “activation time”
  - $\{0, 1\}^{T} \longrightarrow \{0, 1, \ldots, T - 1, \infty\}$
  - $2^{T} \longrightarrow T + 1$
Local Constraints

- seeds: $i$ such that $x_i^0 = 1$, i.e. $t_i = 0$.

- Dynamical rule non-seeds
  \[ t_i = \min \left\{ t : \sum_{j \in \partial i} w_{ji} \mathbb{1} \left[ t_j < t \right] \geq \theta_i \right\} = f \left( \{ t_j \} \right) \]

- Note that
  \[ \mathbb{1} \left[ t_i = f \left( \{ t_j \} \right) \right] = \mathbb{1} \left[ \sum_{j \in \partial i} w_{ji} \mathbb{1} \left[ t_j < t \right] \geq \theta_i \right] \mathbb{1} \left[ \sum_{j \in \partial i} w_{ji} \mathbb{1} \left[ t_j < t - 1 \right] < \theta_i \right] \]

- Local constraint $\Psi_i (t_i) = \mathbb{1} \left[ t_i = 0 \right] + \mathbb{1} \left[ t_i = f \left( \{ t_j \} \right) \right]$
For small $T$ or $\theta$, this is known as the problem of domination:

- $\theta = 1$, $T$ arbitrary: $T$-domination.
- $T = 2$, $\theta$ arbitrary: $\theta$-tuple set domination ($\theta=1$ dominating set)
- Their optimization version is NP-Hard
- Some optima can be proven in lattices, approximation algorithms, etc.
The model

- Interactions are between a site $i$ and its neighbors $\{t_j\}_{j \in \partial i}$

  Beware of **double** interactions ($\Psi_i, \Psi_j$ share both $t_i$ and $t_j$)

- Standard trick: refactor with the *dual* factor graph
  - variables $(t_i^{(ij)}, t_i^{(ij)})$ per edge $(ij)$ (we may keep also single site vars $t_i$)
  - interaction potential $Q_i$ involving variables in all edges of $i$:

  $$ P(t) \propto \prod_i Q_i(t_i) \prod_i \phi_i(t_i) $$

  $$ Q_i(t_i, \{(t_i^{(ik)}, t_k^{(ik)})\}_{k \in \partial i}) = \delta(t_i, f(\{t_k^{(ik)}\})) \prod_{k \in \partial i} \delta(t_i^{(ik)}, t_i) $$

  $$ \phi_i(t_i) = e^{-\mu_i \delta(t_i,0) - \epsilon_i \delta(t_i,\infty)} $$
Dual factor graph

\[ H_{i\ell}(t_i, t_{\ell}) \propto \phi_i(t_i) \sum_{\{t_k, k \in \partial i \setminus \ell\}} \psi_i(t_i) \prod_{k \in \partial i \setminus \ell} H_{ki}(t_k, t_i) \]

\[ \begin{align*} 
\phi_i(t_i) &= e^{-\mu_i \delta(t_i, 0) - \epsilon_i \delta(t_i, \infty)} \\
\mathcal{E}(t) &= \sum_i \mathcal{E}_i(t_i) = \sum_i \epsilon_i \delta(t_i, \infty) + \mu_i \delta(t_i, 0)
\end{align*} \]
BP Equations

\[ m_{ij}(t_i, t_j) = \begin{cases} 
\sum_{\{t_k, k \in \partial i \setminus j\}} \prod_{k \in \partial i \setminus j} m_{ki}(t_k, t_i) - \mu_i & t_i = 0 \\
\sum_{\{t_k, k \in \partial i \setminus j\} \text{ s.t:}} \prod_{k \in \partial i \setminus j} m_{ki}(t_k, t_i) - \epsilon_i & t_i = T \\
\sum_{\{t_k, k \in \partial i \setminus j\} \text{ s.t:}} \prod_{k \in \partial i \setminus j} m_{ki}(t_k, t_i) & \text{else}
\end{cases} \]
For $H_{ij}(t_i, t_j) = \lim_{\beta \to \infty} \frac{1}{\beta} \log m_{ij}(t_i, t_j)$, e.g. for $\mu_i = \beta \mu_i', \epsilon_i = \beta \epsilon_i'$,

$$H_{ij}(t_i, t_j) = \begin{cases} 
\max_{\{t_k, k \in \partial i \setminus j\}} \sum_{k \in \partial i \setminus j} H_{ki}(t_k, 0) - \mu_i' & t_i = 0 \\
\max_{\{t_k, k \in \partial i \setminus j\} \text{ s.t:}} \sum_{\ell \in \partial i} w_{\ell i} \mathbb{1}[t_{\ell} < d - 1] < \theta_i \\
\max_{\{t_k, k \in \partial i \setminus j\} \text{ s.t:}} \sum_{\ell \in \partial i} w_{\ell i} \mathbb{1}[t_{\ell} \leq t_i - 1] \geq \theta_i, \\
\sum_{\ell \in \partial i} w_{\ell i} \mathbb{1}[t_{\ell} < t_i - 1] < \theta_i \\
\sum_{k \in \partial i \setminus j} H_{ki}(t_k, t_i) - \epsilon_i' & t_i = T \\
\max_{\{t_k, k \in \partial i \setminus j\} \text{ s.t:}} \sum_{\ell \in \partial i} w_{\ell i} \mathbb{1}[t_{\ell} < t_i - 1] < \theta_i \\
\sum_{k \in \partial i \setminus j} H_{ki}(t_k, t_i) & \text{else}
\end{cases}$$
Computable equations for $K, \Theta, T >> 1$

Consider the case $0 < t_i < T$:

$$m_{ij}(t_i, t_j) = \sum_{\{t_k, k \in \partial i \setminus j\} \text{ s.t.}} \prod_{k \in \partial i \setminus j} m_{ki}(t_k, t_i)$$

This quantity can obviously be computed efficiently in terms of

$$Q_{ij}(\theta_1, \theta_2, t_i) = \sum_{\{t_k, k \in \partial i \setminus j\} \text{ s.t.}} \prod_{k \in \partial i \setminus j} m_{ki}(t_k, t_i)$$

For fixed $i$, this is the distribution of $\sum_{k \in i \setminus j} (x_k, y_k)$ where

$$P_k(x_k, y_k) = \sum_{t_k} \delta(x_k, w_{ki}[t_k < t_i - 1]) \delta(y_k, w_{ki}[t_k \leq t_i - 1]) m_{ki}(t_k, t_i)$$

i.e. a convolution $P_{k_1} * P_{k_2} * \cdots * P_{k_{K-1}}$ (can be computed naively in time $O(KT\Theta^4)$ or in time $O(KT\Theta^2 \log(\Theta))$ by using FFT).
Directed times, two level optimization

- An alternative model is with directed (cavity) times, $t_{ij}$ (time of activation of $i$ in absence of $j$).
- Two-level optimization
- When $\epsilon = 0$, $P_{ij}(t_{ij}, t_{ji}) = P_{ij}(t_{ij})$, and this becomes equivalent to the formulation in [Otha&Sasa EPL2010].
- Positive $\epsilon$ correlates $t_{ij}$ and $t_{ji}$. 

movie?
Some results for random regular graphs

On RRG of degree $K$, one may assume a single link representation in the limit $N \to \infty$:

$$H(s, t) \propto \sum_{n_0+n_++n_-=K-1} \frac{(K-1)!}{n_0!n_-!n_+!} p_t^{K-1-n_-} m_t^{n_-} H(t-1, t)^{n_0}$$

$$H(0, s) \propto e^{-\beta \mu} p_0^{K-1}$$

$$H(\infty, s) \propto e^{-\beta \epsilon} \sum_{n_- \leq \theta - 1 - 1[s < T]} \binom{K-1}{n_-} (H(T, \infty) + H(\infty, \infty))^{K-1-n_-} m_\infty^{n_-}$$

with cumulative messages $p_t = \sum_{s \geq t} H(s, t)$, $m_t = \sum_{s < t-1} H(s, t)$

(next talk on random graphs by Luca Dall’Asta)
Activation transitions

\[ \theta = 2, K = 3 \] continuous transition for \( \varepsilon = 0 \), discontinuous for e.g. \( \varepsilon = 0.5 \) (optimization-induced discontinuity)
rare trajectories

c=3, k=2

N=100 binomial
N=100 BP
N=50 binomial
N=50 BP

fraction of seeds

P(percolation)
Other methods

- Simulated annealing: good in some cases, much slower in large instances

\[ K = 5, \theta = 4, \Delta \rho_0 = \frac{\rho_{0}^{SA}}{\rho_{0}^{MS}} - 1 \]

- LP relaxation / CPLEX: seems unfeasible even for small sizes (\(N \sim 100\)).

- Each SA run with \(M = 10^5\), \(N = 1000\) was around 73 hours. (MS = 10m, linear in \(N\))
Applications

- Viral marketing / influence optimization
  - e-pinions (405740 directed edges, 75879 vertices)
    - With $\theta_i = \left\lceil \frac{|\partial i|}{2} \right\rceil + 1$, you can activate 50% with 1100 seeds
  - twitter, facebook (huge! more than $10^8$ vertices, $10^{10}$ edges but good data hard to get)

- Default cascades in finance (or how to block them cheaply)

- Analysis of signaling cascades in cells

- Not many alternatives for global optimization. Note that even Brandes’ inbetweeness centrality algorithm is $|E| \cdot |V|$ (compared with $|E| \cdot T\Theta^2$ (per iteration) of BP/MS)
N=75879
M=405740
(16000 zero out-degree)
\[
\theta_i = \left\lfloor \frac{|\partial_i|}{2} \right\rfloor + 1
\]

\[
\text{cost} = |E| T \Theta^2 \quad \text{cost inbetweeness algorithm} = |E| \cdot |V|
\]
**Stochastic model: independent cascade dynamics**

stage 1: t=0, seeds extracted, stage 2: \( t_{ij} \) are drawn independently at random with \( p_{ij} \)

\[
x^T_j = \begin{cases} 
1 & \text{if } x^T_{j-1} = 1 \\
t_{ij} & \text{otherwise}
\end{cases} \quad (\forall i \in A(\tau-1), \forall j \in \partial i : x^T_{j-1} = 0)
\]

\[
E(x^0, t) = \mu \sum_i c_i x_i^0 - \sum_i x_i^T(x^0, t)
\]

\[
\bar{x}^0 = \arg \min_{x^0} \mathbb{E}_t E(x^0, t)
\]

**two-stage stochastic optimization problem**

Blocking activation processes in the stochastic model

\[ x^0_i = s_i t_i \]

\[ t_i \in \{0, 1\} \text{ with } P[t_i = 1] = p_i \text{ (random background at } t=0) \]

\[ s_i \in \{0, 1\} \quad s_i = 0 \rightarrow \text{ cannot be a seed} \]

\[ \mathcal{E}(s, t) = \mu \sum_i (1 - s_i) c_i + \sum_i x^T_i(s, t) \]

\[ \bar{x}^0 = \arg \min_{x^0} \mathbb{E}_t \mathcal{E}(x^0, t) \]

minimize final activity

very interesting connection with SIR model