Bayesian inference of cascades on networks

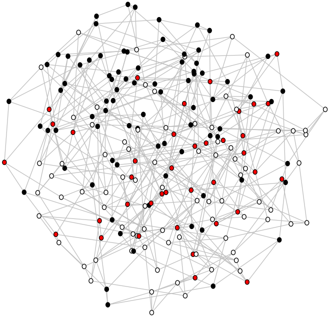
Alfredo Braunstein

Politecnico di Torino

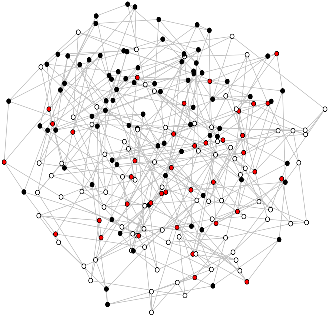
NIPS
Montreal, Dec 2015
A. Ingrosso, J. Bindi, L. Dall'Asta

The *patient zero* or *index case* problem

INPUT



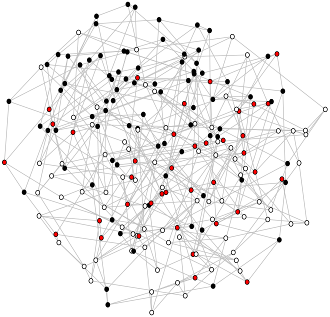
The *patient zero* or *index case* problem



INPUT

- A contact network in a community:
  - Hospital wards [**Vanhems'13**]
  - Livestock Surveillance [**Bajardi'12**]
  - Many others, e.g.: Sexual contacts [**Rocha'10**], Proximity in a closed environment [**Isella'10**]

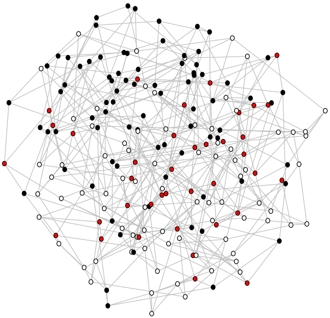
The *patient zero* or *index case* problem



INPUT

- A contact network in a community:
  - Hospital wards [**Vanhems'13**]
  - Livestock Surveillance [**Bajardi'12**]
  - Many others, e.g.: Sexual contacts [**Rocha'10**], Proximity in a closed environment [**Isella'10**]
- An epidemic *snapshot* at time $t = T$
  - ○ Susceptible
  - ● Infected
  - ● Recovered

The *patient zero* or *index case* problem



INPUT

- A contact network in a community:
  - Hospital wards [**Vanhems'13**]
  - Livestock Surveillance [**Bajardi'12**]
  - Many others, e.g.: Sexual contacts [**Rocha'10**], Proximity in a closed environment [**Isella'10**]
- An epidemic *snapshot* at time $t = T$
  - ○ Susceptible
  - ● Infected
  - ● Recovered

OUTPUT

- Find the ***source*** node at time $t = 0$

Related Problems

INPUTS

- Various types of observations: time and space scattered and noisy
- Unknown epidemic "age" T
- Time-evolving networks
- Multiple sources

Related Problems

INPUTS

- Various types of observations: time and space scattered and noisy
- Unknown epidemic "age" T
- Time-evolving networks
- Multiple sources

OUTPUTS

- Identifying contagion paths and undiscovered positives
- Predicting of future development of an outbreak
- Reconstructing the contact network (from the observation of multiple cascades)

The (discrete) SIR process on a network

Per-vertex variables $x_i \in \{\mathbb{S}, \mathbf{I}, \mathbf{R}\}$. At each t, each **infected** node $x_i^t \in \mathbf{I}$

- attempts **contagion** to susceptible neighbors in $x_j^t \in \mathbb{S}$ with probability λ. If successful, $x_j^{t+1} = \mathbf{I}$
- attempts **recovery** with probability μ. If successful, $x_i^{t+1} = \mathbf{R}$

The (discrete) SIR process on a network

Per-vertex variables $x_i \in \{\mathbb{S}, \mathbf{I}, \mathbf{R}\}$. At each t, each **infected** node $x_i^t \in \mathbf{I}$

- attempts **contagion** to susceptible neighbors in $x_j^t \in \mathbb{S}$ with probability λ. If successful, $x_j^{t+1} = \mathbf{I}$
- attempts **recovery** with probability μ. If successful, $x_i^{t+1} = \mathbf{R}$

SIR Markov Chain

$$P\left(\mathbf{x}^{t+1}|\mathbf{x}^t\right) = \prod_i P\left(x_i^{t+1}|\mathbf{x}^t\right),$$

$$P(x_i^{t+1} = \mathbb{S}|\mathbf{x}^t) = \mathbb{I}[x_i^t = \mathbb{S}] \prod_{j\in\partial i} (1 - \lambda \mathbb{I}[x_j^t = \mathbf{I}])$$

$$P(x_i^{t+1} = \mathbf{I}|\mathbf{x}^t) = \mathbb{I}[x_i^t = \mathbf{I}](1-\mu) + \mathbb{I}[x_i^t = \mathbb{S}](1 - \prod_{j\in\partial i} (1 - \lambda \mathbb{I}[x_j^t = \mathbf{I}]))$$

$$P(x_i^{t+1} = \mathbf{R}|\mathbf{x}^t) = \mathbb{I}[x_i^t = \mathbf{I}]\mu + \mathbb{I}[x_i^t = \mathbf{R}]$$

Approaches

- Topological centrality measures **[Shah'10], [Comin'11], [Zhu'12]**

Approaches

- Topological centrality measures **[Shah'10], [Comin'11], [Zhu'12]**
- Bayesian inference: compute $P\left(\mathbf{x}^0|\mathbf{x}^T\right)$
  - "Brute-Force" Monte Carlo (variant: use soft compatibility [**Antulov-Fantulin'14**])
  - *Naive Bayes*
  - **Belief Propagation**

Naive Bayes (1/3)

- Assume the following naive MF structure for the distribution $P\left(\mathbf{x}^T|\mathbf{x}^0\right) \simeq \prod_i P\left(x_i^T|\mathbf{x}^0\right)$
- Marginals $P\left(x_i^T|\mathbf{x}^0\right)$ can be computed either with MC or with Dynamical Message-Passing **[Lokhov, Mézard & al.'14]**

Naive Bayes (1/3)

- Assume the following naive MF structure for the distribution $P\left(\mathbf{x}^T|\mathbf{x}^0\right) \simeq \prod_i P\left(x_i^T|\mathbf{x}^0\right)$
- Marginals $P\left(x_i^T|\mathbf{x}^0\right)$ can be computed either with MC or with Dynamical Message-Passing **[Lokhov, Mézard & al.'14]**
- Then maximize over $\mathbf{x}^0$ the likelihood $P\left(\mathscr{O}|\mathbf{x}^0\right) \simeq \prod_i P\left(x_i^T = \mathscr{O}_i|\mathbf{x}^0\right)$

Naive Bayes (1/3)

- Assume the following naive MF structure for the distribution $P\left(\mathbf{x}^T|\mathbf{x}^0\right)\simeq\prod_i P\left(x_i^T|\mathbf{x}^0\right)$
- Marginals $P\left(x_i^T|\mathbf{x}^0\right)$ can be computed either with MC or with Dynamical Message-Passing **[Lokhov, Mézard & al.'14]**
- Then maximize over $\mathbf{x}^0$ the likelihood $P\left(\mathscr{O}|\mathbf{x}^0\right)\simeq\prod_i P\left(x_i^T=\mathscr{O}_i|\mathbf{x}^0\right)$
- Is the approximation accurate? If no, is it because of Naive MF?

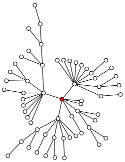
Naive Bayes (1/3)

- Assume the following naive MF structure for the distribution $P\left(\mathbf{x}^T|\mathbf{x}^0\right) \simeq \prod_i P\left(x_i^T|\mathbf{x}^0\right)$
- Marginals $P\left(x_i^T|\mathbf{x}^0\right)$ can be computed either with MC or with Dynamical Message-Passing **[Lokhov, Mézard & al.'14]**
- Then maximize over $\mathbf{x}^0$ the likelihood $P\left(\mathscr{O}|\mathbf{x}^0\right) \simeq \prod_i P\left(x_i^T = \mathscr{O}_i|\mathbf{x}^0\right)$
- Is the approximation accurate? If no, is it because of Naive MF?
- Note that Naive MF can easily be replaced by e.g.:
  $P\left(\mathbf{x}^T|\mathbf{x}^0\right) \simeq \prod_{\langle ij\rangle} \frac{P\left(x_i^T,x_j^T|\mathbf{x}^0\right)}{P\left(x_i^T|\mathbf{x}^0\right)P\left(x_j^T|\mathbf{x}^0\right)} \prod_i P\left(x_i^T|\mathbf{x}^0\right)$ **[Lokhov, Mézard & al.'14,** maybe**]**

Naive Bayes (1/3)

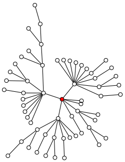
- Assume the following naive MF structure for the distribution $P\left(\mathbf{x}^T|\mathbf{x}^0\right) \simeq \prod_i P\left(x_i^T|\mathbf{x}^0\right)$
- Marginals $P\left(x_i^T|\mathbf{x}^0\right)$ can be computed either with MC or with Dynamical Message-Passing **[Lokhov, Mézard & al.'14]**
- Then maximize over $\mathbf{x}^0$ the likelihood $P\left(\mathcal{O}|\mathbf{x}^0\right) \simeq \prod_i P\left(x_i^T = \mathcal{O}_i|\mathbf{x}^0\right)$
- Is the approximation accurate? If no, is it because of Naive MF?
- Note that Naive MF can easily be replaced by e.g.:
  $P\left(\mathbf{x}^T|\mathbf{x}^0\right) \simeq \prod_{\langle ij \rangle} \frac{P\left(x_i^T, x_j^T|\mathbf{x}^0\right)}{P\left(x_i^T|\mathbf{x}^0\right)P\left(x_j^T|\mathbf{x}^0\right)} \prod_i P\left(x_i^T|\mathbf{x}^0\right)$ **[Lokhov, Mézard & al.'14,** maybe**]**

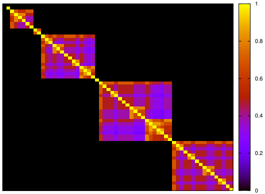
Naive Bayes (2/3)

Graph	Ising	SI
$i = \bullet$	$Pearson(\sigma_j,\sigma_k\|\sigma_i=1)$	$Pearson(x_j^T,x_k^T\|x_i^T=I)$
		

Naive Bayes (2/3)

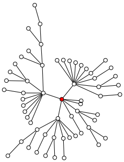
Graph	Ising	SI
$i =$ ●	$Pearson(\sigma_j,\sigma_k \mid \sigma_i=1)$	$Pearson(x_j^T, x_k^T \mid x_i^T = I)$

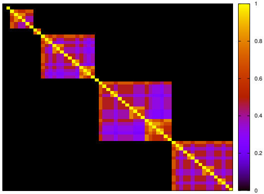


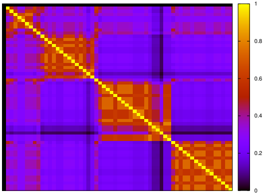


Naive Bayes (2/3)

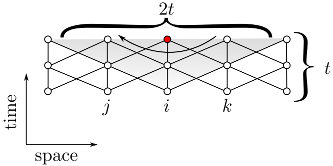
Graph	Ising	SI
$i =$ ●	$Pearson(\sigma_j, \sigma_k \mid \sigma_i = 1)$	$Pearson(x_j^T, x_k^T \mid x_i^T = I)$





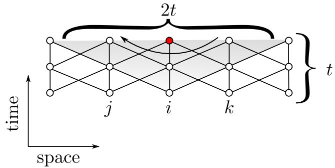


Naive Bayes (3/3)



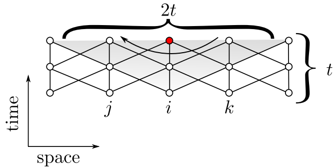
- Sites x_j^T and x_k^T interact e.g. through x_i^{T-1}

Naive Bayes (3/3)



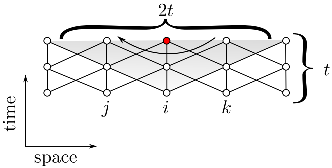
- Sites x_j^T and x_k^T interact e.g. through x_i^{T-1}
- Interactions between surface ($t = T$) variables are long-range

Naive Bayes (3/3)



- Sites x_j^T and x_k^T interact e.g. through x_i^{T-1}
- Interactions between surface ($t = T$) variables are long-range
- The real problem is not to compute $P\left(x_i^T, x_j^T|\mathbf{x}^0\right)$ accurately but to give a "functional" parametrization of $P\left(\mathbf{x}^T|\mathbf{x}^0\right)$

Naive Bayes (3/3)



- Sites x_j^T and x_k^T interact e.g. through x_i^{T-1}
- Interactions between surface ($t = T$) variables are long-range
- The real problem is not to compute $P\left(x_i^T, x_j^T|\mathbf{x}^0\right)$ accurately but to give a "functional" parametrization of $P\left(\mathbf{x}^T|\mathbf{x}^0\right)$
- [Note: to recover the MRF independence property one should fix full columns/trajectories $\mathbf{x}_i^{0:T}$]

Parametrization of trajectories

- We will assume for simplicity $\mu = 1$. The case $\mu < 1$ is similar.

Parametrization of trajectories

- We will assume for simplicity $\mu = 1$. The case $\mu < 1$ is similar.
- Single site trajectories, e.g. $\mathbf{x}_i = \mathbb{SSS}\mathbf{IRRRRR}$ can be parametrized by a single *infection time* t_i

Parametrization of trajectories

- We will assume for simplicity $\mu = 1$. The case $\mu < 1$ is similar.
- Single site trajectories, e.g. $\mathbf{x}_i = \mathbb{SSS}$**IRRRRR** can be parametrized by a single *infection time* t_i
- We can divide the process in two parts:

Parametrization of trajectories

- We will assume for simplicity $\mu = 1$. The case $\mu < 1$ is similar.
- Single site trajectories, e.g. $\mathbf{x}_i = \mathbb{SSS}\mathbf{IRRRRR}$ can be parametrized by a single *infection time* t_i
- We can divide the process in two parts:
  1. First, **stochastic** "delays" $s_{ij} \in \{0, \infty\}$ for all $(ij) \in E$ are extracted independently with probabilities $P\left(s_{ij} = 0\right) = \lambda$

Parametrization of trajectories

- We will assume for simplicity $\mu = 1$. The case $\mu < 1$ is similar.
- Single site trajectories, e.g. $\mathbf{x}_i = \mathbb{SSS}$**IRRRRR** can be parametrized by a single *infection time* t_i
- We can divide the process in two parts:
  1. First, **stochastic** "delays" $s_{ij} \in \{0, \infty\}$ for all $(ij) \in E$ are extracted independently with probabilities $P\left(s_{ij} = 0\right) = \lambda$
  1. Afterwards, all $t_i \neq 0$ can be computed **deterministically** with the following self-consisting equations

$$t_i = 1 + \min_{j \in \partial i} \left\{ t_j + s_{ji} \right\}$$

Parametrization of trajectories

- We will assume for simplicity $\mu = 1$. The case $\mu < 1$ is similar.
- Single site trajectories, e.g. $\mathbf{x}_i = \mathbb{SSS}$**IRRRRR** can be parametrized by a single *infection time* t_i
- We can divide the process in two parts:
  1. First, **stochastic** "delays" $s_{ij} \in \{0, \infty\}$ for all $(ij) \in E$ are extracted independently with probabilities $P\left(s_{ij} = 0\right) = \lambda$
  1. Afterwards, all $t_i \neq 0$ can be computed **deterministically** with the following self-consisting equations

$$t_i = 1 + \min_{j \in \partial i} \left\{ t_j + s_{ji} \right\}$$

- Key: stochastic parameters are *independent*

A factorized distribution

$$\mathscr{P}\left(\mathbf{t}|\mathbf{x}^{0}\right)=\sum_{\mathbf{s}}\mathscr{P}\left(\mathbf{t}|\mathbf{s},\mathbf{x}^{0}\right)\mathscr{P}\left(\mathbf{s}\right)$$

A factorized distribution

$$\mathscr{P}\left(\mathbf{t}|\mathbf{x}^0\right)=\sum_{\mathbf{s}}\mathscr{P}\left(\mathbf{t}|\mathbf{s},\mathbf{x}^0\right)\mathscr{P}\left(\mathbf{s}\right)$$

Define

- $\omega_{ij}\left(s_{ij}\right)=\lambda\delta\left(s_{ij},0\right)+\left(1-\lambda\right)\delta\left(s_{ij},\infty\right)$
- $\phi_i\left(t_i,\mathbf{t}_{\partial i},\mathbf{s}_{\partial i},x_i^0\right)=\delta\left(t_i,\delta\left(x_i^0;\mathbf{I}\right)\left(1+\min_{j\in\partial i}\left\{t_j+s_{ji}\right\}\right)\right)$

A factorized distribution

$$\mathscr{P}\left(\mathbf{t}|\mathbf{x}^0\right)=\sum_{\mathbf{s}}\mathscr{P}\left(\mathbf{t}|\mathbf{s},\mathbf{x}^0\right)\mathscr{P}\left(\mathbf{s}\right)$$

Define

- $\omega_{ij}\left(s_{ij}\right)=\lambda\delta\left(s_{ij},0\right)+\left(1-\lambda\right)\delta\left(s_{ij},\infty\right)$
- $\phi_i\left(t_i,\mathbf{t}_{\partial i},\mathbf{s}_{\partial i},x_i^0\right)=\delta\left(t_i,\delta\left(x_i^0;\mathbf{I}\right)\left(1+\min_{j\in\partial i}\left\{t_j+s_{ji}\right\}\right)\right)$

$$\mathscr{Q}=\frac{1}{Z}\prod_i\phi_i\prod_{i,j}\omega_{ij}$$

A factorized distribution

$$\mathscr{P}\left(\mathbf{t}|\mathbf{x}^0\right) = \sum_{\mathbf{s}} \mathscr{P}\left(\mathbf{t}|\mathbf{s},\mathbf{x}^0\right)\mathscr{P}\left(\mathbf{s}\right)$$

Define

- $\omega_{ij}\left(s_{ij}\right) = \lambda\delta\left(s_{ij},0\right) + (1-\lambda)\,\delta\left(s_{ij},\infty\right)$
- $\phi_i\left(t_i,\mathbf{t}_{\partial i},\mathbf{s}_{\partial i},x_i^0\right) = \delta\left(t_i,\delta\left(x_i^0;\mathbf{I}\right)\left(1+\min_{j\in\partial i}\left\{t_j+s_{ji}\right\}\right)\right)$

$$\mathscr{Q} = \frac{1}{Z}\prod_i \phi_i \prod_{i,j}\omega_{ij}$$

Then $\mathscr{P}\left(\mathbf{t}|\mathbf{x}^0\right) = \sum_{\mathbf{s}} \mathscr{Q}\left(\mathbf{t},\mathbf{s},\mathbf{x}^0\right)$

Adding priors

- $\mathbf{x}^T$ depends **deterministically** on $\mathbf{t}$: $P\left(\mathbf{x}^T|\mathbf{t}\right) = \prod_i \xi_i\left(t_i, x_i^T\right)$ where $\xi_i\left(t_i, x_i^T\right)$ is the indicator function of

$$\left(x_i^T = \mathbb{S}, t_i > T\right) \vee \left(x_i^T = \mathbf{I}, t_i = T\right) \vee \left(x_i^T = \mathbf{R}, t_i < T\right)$$

Adding priors

- $\mathbf{x}^T$ depends **deterministically** on $\mathbf{t}$: $P\left(\mathbf{x}^T|\mathbf{t}\right)=\prod_i \xi_i\left(t_i,x_i^T\right)$ where $\xi_i\left(t_i,x_i^T\right)$ is the indicator function of

$$\left(x_i^T=\mathbb{S},t_i>T\right)\vee\left(x_i^T=\mathbf{I},t_i=T\right)\vee\left(x_i^T=\mathbf{R},t_i<T\right)$$

- $\mathbf{x}^0$ have a prior concentrated on single-seed initial conditions: $P\left(\mathbf{x}^0\right)=\prod_i \gamma_i\left(x_i^0\right)$ with $\gamma_i\left(x_i^0=I\right)$ very small.

Adding priors

- $\mathbf{x}^T$ depends **deterministically** on $\mathbf{t}$: $P\left(\mathbf{x}^T|\mathbf{t}\right) = \prod_i \xi_i\left(t_i, x_i^T\right)$ where $\xi_i\left(t_i, x_i^T\right)$ is the indicator function of

$$\left(x_i^T = \mathbb{S}, t_i > T\right) \vee \left(x_i^T = \mathbf{I}, t_i = T\right) \vee \left(x_i^T = \mathbf{R}, t_i < T\right)$$

- $\mathbf{x}^0$ have a prior concentrated on single-seed initial conditions: $P\left(\mathbf{x}^0\right) = \prod_i \gamma_i\left(x_i^0\right)$ with $\gamma_i\left(x_i^0 = I\right)$ very small.
- Finally, we can write the **posterior** distribution $P\left(\mathbf{x}^0|\mathbf{x}^T\right) \propto \sum_{\mathbf{t}} P\left(\mathbf{x}^T|\mathbf{t}\right) P\left(\mathbf{t}|\mathbf{x}^0\right) P\left(\mathbf{x}^0\right)$ as

$$P\left(\mathbf{x}^0|\mathbf{x}^T\right) \propto \sum_{\mathbf{t}} \sum_{\mathbf{s}} \prod_{ij} \phi_{ij} \prod_i \phi_i \xi_i \gamma_i \quad (1)$$

Belief Propagation

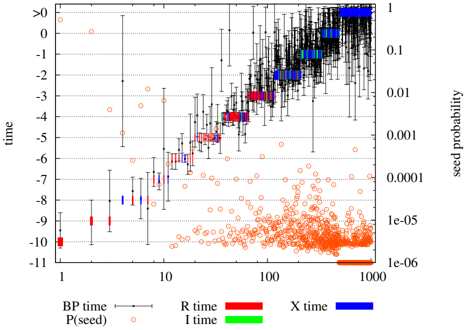
$$P\left(\mathbf{x}^{0}|\mathbf{x}^{T}\right) \propto \sum_{\mathbf{t}}\sum_{\mathbf{s}}\left[\prod_{ij}\phi_{ij}\prod_{i}\phi_{i}\xi_{i}\gamma_{i}\right] = \sum_{\mathbf{t}}\sum_{\mathbf{s}} Q\left(\mathbf{x}^{0},\mathbf{t},\mathbf{s}\right)$$

Single-instance RS cavity equations / Belief Propagation

- Fixed-point equation $\mathbf{m} = F_{BP}\left(\mathbf{m}\right)$ for a vector $\mathbf{m}$ (called *cavity marginals* or *messages*) that is solved by iteration.
  - On a fixed point (approximate) marginals $P\left(t_i|\mathbf{x}^{T}\right)$ or $P\left(x_i^{0}|\mathbf{x}^{T}\right)$ can be computed.
  - **Fast**: each iteration is often linear in the number of edges, needed number of iterations is usually logarithmic
  - Exact if the ***factor graph*** is acyclic

Results on random graphs

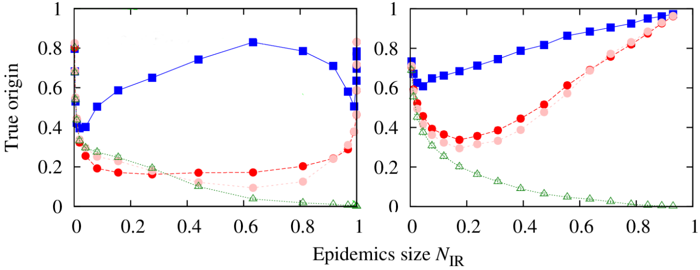
$$N = 1000,\ k = 4,\ \lambda = 0.5,\ \mu = 0.5,\ \gamma = 10^{-6}$$



unknown $T - t_0 = 10$, 60% observed nodes

Results on random graphs

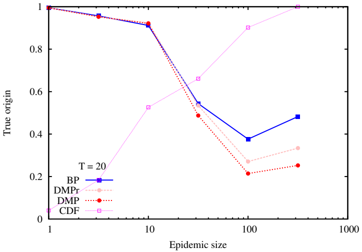
RRG $N = 1000, k = 4, \mu = 0.5,\ T - t_0 = 10$ and preferential attachment $\langle k \rangle = 4,\ N = 1000,\ T - t_0 = 5$



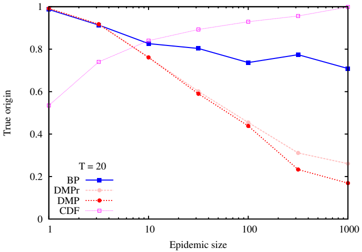
- Belief Propagation
- Dynamic message-passing **[Lokhov, Mézard, Ohta & Zdeborová'14]**
- Jordan centrality **[Zhu & Ying'12]**

Time-evolving networks

- Temporal networks can be analyzed by using a modified ω_{ij}



proximity [**Isella et al.'10**]



sexual [**Rocha et al'10**]

Inference of parameters

- The likelihood of λ, μ can be computed as:
  $P\left(\mathbf{x}^T|\lambda,\mu\right) = \sum_{t,\mathbf{g},\mathbf{x}^0} P\left(\mathbf{x}^T|\mathbf{t},\mathbf{g}\right) P\left(\mathbf{t},\mathbf{g}|\mathbf{x}^0,\lambda,\mu\right) P\left(\mathbf{x}^0\right)$

Inference of parameters

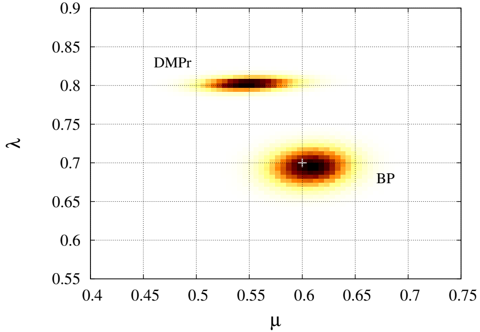
- The likelihood of λ,μ can be computed as:
$P\left(\mathbf{x}^{T}|\lambda,\mu\right)=\sum_{t,\mathbf{g},\mathbf{x}^{0}} P\left(\mathbf{x}^{T}|\mathbf{t},\mathbf{g}\right) P\left(\mathbf{t},\mathbf{g}|\mathbf{x}^{0},\lambda,\mu\right) P\left(\mathbf{x}^{0}\right) = Z$

Inference of parameters

- The likelihood of λ, μ can be computed as:
$P\left(\mathbf{x}^{T}|\lambda,\mu\right) = \sum_{t,\mathbf{g},\mathbf{x}^{0}} P\left(\mathbf{x}^{T}|\mathbf{t},\mathbf{g}\right) P\left(\mathbf{t},\mathbf{g}|\mathbf{x}^{0},\lambda,\mu\right) P\left(\mathbf{x}^{0}\right) = Z \simeq Z_{Bethe}$

Inference of parameters

- The likelihood of λ,μ can be computed as:
  $P(\mathbf{x}^T|\lambda,\mu) = \sum_{t,\mathbf{g},\mathbf{x}^0} P(\mathbf{x}^T|\mathbf{t},\mathbf{g})\, P(\mathbf{t},\mathbf{g}|\mathbf{x}^0,\lambda,\mu)\, P(\mathbf{x}^0) = Z \simeq Z_{Bethe}$



RRG $k = 4, N = 1000$

Interleaved BP+GA

- We need to maximize the log-likelihood $\mathscr{L} = \log Z \simeq -f_{Bethe}$ with respect to λ (and/or μ), but

$$\frac{\partial}{\partial\lambda}\left[f\left(\mathbf{m},\lambda\right)\right] = \nabla_{\mathbf{m}} f \cdot \frac{\partial \mathbf{m}}{\partial\lambda} + \frac{\partial f}{\partial\lambda}$$

Interleaved BP+GA

- We need to maximize the log-likelihood $\mathcal{L} = \log Z \simeq -f_{Bethe}$ with respect to λ (and/or μ), but

$$\frac{\partial}{\partial \lambda}\left[f\left(\mathbf{m},\lambda\right)\right] = \nabla_{\mathbf{m}} f \cdot \frac{\partial \mathbf{m}}{\partial \lambda} + \frac{\partial f}{\partial \lambda} = \frac{\partial f}{\partial \lambda}$$

because $\nabla_{\mathbf{m}} f \equiv 0$ on a FP of BP, as the BP solution is a variational critical point of f.

Interleaved BP+GA

- We need to maximize the log-likelihood $\mathscr{L} = \log Z \simeq -f_{Bethe}$ with respect to λ (and/or μ), but

$$\frac{\partial}{\partial\lambda}\left[f\left(\mathbf{m},\lambda\right)\right] = \nabla_{\mathbf{m}} f \cdot \frac{\partial \mathbf{m}}{\partial\lambda} + \frac{\partial f}{\partial\lambda} = \frac{\partial f}{\partial\lambda}$$

  because $\nabla_{\mathbf{m}} f \equiv 0$ on a FP of BP, as the BP solution is a variational critical point of f.

- Now $\frac{\partial f}{\partial\lambda}(\mathbf{m},\lambda) = -\frac{1}{Z}\frac{\partial}{\partial\lambda}\left\{\sum_{\mathbf{t},\mathbf{s}} e^{\sum_i \log\psi_i + \sum_{\langle ij\rangle}\log\phi_{ij}}\right\} = -\sum_{\mathbf{t},\mathbf{s}}\sum_{\langle ij\rangle}\left\{\frac{\partial}{\partial\lambda}\log\phi_{ij}\right\}\frac{1}{Z}e^{\sum_i \log\psi_i + \sum_{\langle ij\rangle}\log\phi_{ij}} = -\sum_{\langle ij\rangle}\left\langle\frac{\partial}{\partial\lambda}\log\phi_{ij}\right\rangle$, i.e. the computation of an observable

Interleaved BP+GA

- We need to maximize the log-likelihood $\mathscr{L} = \log Z \simeq -f_{Bethe}$ with respect to λ (and/or μ), but

$$\frac{\partial}{\partial\lambda}\left[f\left(\mathbf{m},\lambda\right)\right] = \nabla_{\mathbf{m}} f \cdot \frac{\partial \mathbf{m}}{\partial\lambda} + \frac{\partial f}{\partial\lambda} = \frac{\partial f}{\partial\lambda}$$

  because $\nabla_{\mathbf{m}} f \equiv 0$ on a FP of BP, as the BP solution is a variational critical point of f.
- Now $\frac{\partial f}{\partial\lambda}\left(\mathbf{m},\lambda\right) = -\frac{1}{Z}\frac{\partial}{\partial\lambda}\left\{\sum_{\mathbf{t},\mathbf{s}} e^{\sum_i \log\psi_i + \sum_{\langle ij\rangle}\log\phi_{ij}}\right\} = -\sum_{\mathbf{t},\mathbf{s}}\sum_{\langle ij\rangle}\left\{\frac{\partial}{\partial\lambda}\log\phi_{ij}\right\}\frac{1}{Z}e^{\sum_i \log\psi_i + \sum_{\langle ij\rangle}\log\phi_{ij}} = -\sum_{\langle ij\rangle}\left\langle\frac{\partial}{\partial\lambda}\log\phi_{ij}\right\rangle$, i.e. the computation of an observable
- Gradient updates can be interleaved with BP updates to recover the parameters in one single convergence
- Same fixed points as EM but faster

Inference of network topology

- The same approach can be used to infer single-link parameters from multiple cascades:

$$\frac{\partial \log\left(\prod_{\mu=1}^{M} Z^{\mu}\right)}{\partial \lambda_{ij}} = -\sum_{\mu=1}^{M} \frac{\partial f^{\mu}}{\partial \lambda_{ij}} = \sum_{\mu=1}^{M} \left\langle \frac{\partial}{\partial \lambda_{ij}} \log \phi_{ij} \right\rangle_{\mu}$$

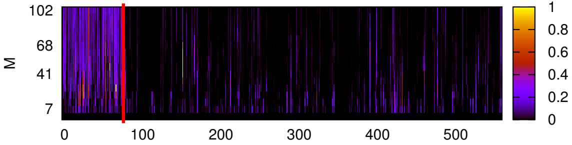
- The factor graph consists in M **independent** (fully-connected $N \times N$) networks that share the matrix λ

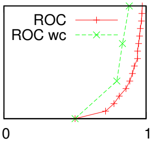
Inference of network topology

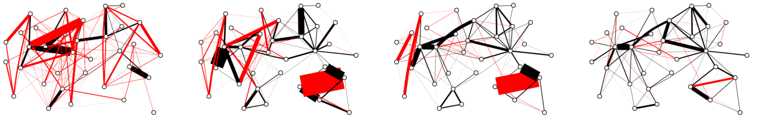
Karate club network ($N = 34, \lambda = 0.3, \mu = 0.4, T = 5$)

Inference of network topology

Karate club network $(N = 34, \lambda = 0.3, \mu = 0.4, T = 5)$







$M = 7$ $M = 41$ $M = 68$ $M = 102$

- ROC area with $N(N-1)/2$ points using sorted inferred values λ_{ij}

Conclusions

- The Bethe parametrization of the probability space of dynamical trajectories gives great **flexibility!**
  - Gives a practical solution to the patient-zero problem on real and synthetic networks (exact on acyclic graphs) with many types of observations (incomplete, noisy, etc)
  - Allows to tackle the problem of inferring edges (ij) in the supporting network having **no direct access to co-infection events** $x_j^{t-1} = \mathbb{I}, x_i^{t-1} = \mathbb{S}$ and $x_i^t = \mathbb{I}$.
  - More on: PRL 112(11) 118701 (2014), JSTAT (10), P10016 (2014)

Thank you!