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**

Alfredo Braunstein
Bayesian inference of cascades on networks
Politecnico di Torino
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
The patient zero or index case problem

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 $x_i \in \{S, I, R\}$. At each $t$, each \textbf{infected} node $x_i^t \in I$

- attempts \textbf{contagion} to susceptible neighbors in $x_j^t \in S$ with probability $\lambda$. If successful, $x_j^{t+1} = I$

- attempts \textbf{recovery} with probability $\mu$. If successful, $x_i^{t+1} = R$
The (discrete) SIR process on a network

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

- attempts **contagion** to susceptible neighbors in $x_j^t \in S$ with probability $\lambda$. If successful, $x_j^{t+1} = I$

- attempts **recovery** with probability $\mu$. If successful, $x_i^{t+1} = R$

\[
P(x_i^{t+1}|x^t) = \prod_i P(x_i^{t+1}|x^t),
\]

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

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

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

- Topological centrality measures [Shah’10], [Comin’11], [Zhu’12]
The problem and classical approaches

Approaches

- Topological centrality measures [Shah’10], [Comin’11], [Zhu’12]
- Bayesian inference: compute $P(x^0|x^T)$
  - “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

$$ P(x^T | x^0) \approx \prod_i P(x_i^T | x^0) $$

- Marginals $$ P(x_i^T | x^0) $$ 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
  \[ P(x^T|x^0) \approx \prod_i P(x_i^T|x^0) \]

- Marginals \( P(x_i^T|x^0) \) can be computed either with MC or with Dynamical Message-Passing [Lokhov, Mézard & al.’14]

- Then maximize over \( x^0 \) the likelihood \( P(\mathcal{O}|x^0) \approx \prod_i P(x_i^T = \mathcal{O}_i|x^0) \)
The problem and classical approaches

Naive Bayes (1/3)

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

- Marginals \( P(x_i^T | x^0) \) can be computed either with MC or with Dynamical Message-Passing [Lokhov, Mézard & al.'14]

- Then maximize over \( x^0 \) the likelihood \( P(O | x^0) \approx \prod_i P(x_i^T = O_i | x^0) \)

- Is the approximation accurate? If no, is it because of Naive MF?

- Note that Naive MF can easily be replaced by e.g.:
  \[ P(x^T | x^0) \approx \prod_{\langle ij \rangle} \frac{P(x_i^T, x_j^T | x^0)}{P(x_i^T | x^0)P(x_j^T | x^0)} \prod_i P(x_i^T | x^0) \] [Lokhov, Mézard & al.'14, maybe]
Naive Bayes (1/3)

- Assume the following naive MF structure for the distribution
  \[ P(x^T | x^0) \approx \prod_i P(x_i^T | x^0) \]
- Marginals \[ P(x_i^T | x^0) \] can be computed either with MC or with
  Dynamical Message-Passing \textbf{[Lokhov, Mézard & al.'14]}
- Then maximize over \( x^0 \) the likelihood \[ P(\mathcal{O} | x^0) \approx \prod_i P(x_i^T = \mathcal{O}_i | x^0) \]
- Is the approximation accurate? If no, is it because of Naive MF?
- Note that Naive MF can easily be replaced by e.g.:
  \[ P(x^T | x^0) \approx \prod_{\langle ij \rangle} \frac{P(x_i^T, x_j^T | x^0)}{P(x_i^T | x^0)P(x_j^T | x^0)} \prod_i P(x_i^T | x^0) \textbf{[Lokhov, Mézard & al.'14, maybe]} \]
Naive Bayes (2/3)

Graph

Ising

\[ \text{Pearson}(\sigma_j, \sigma_k | \sigma_i = 1) \]

SI

\[ \text{Pearson}(x_j^T, x_k^T | x_i^T = 1) \]
The problem and classical approaches

Naive Bayes (2/3)

Graph

\[ i = \bullet \]

Ising

\[ \text{Pearson}(\sigma_j, \sigma_k | \sigma_i = 1) \]

SI

\[ \text{Pearson}(x_j^T, x_k^T | x_i^T = 1) \]
The problem and classical approaches

Naive Bayes (2/3)

Graph

\[
i = \bullet
\]

Ising

\[
\text{Pearson}(\sigma_j, \sigma_k | \sigma_i = 1)
\]

SI

\[
\text{Pearson}(x_j^T, x_k^T | x_i^T = 1)
\]
The problem and classical approaches

Naive Bayes (3/3)

- Sites $x_j^T$ and $x_k^T$ interact e.g. through $x_i^{T-1}$
The problem and classical approaches

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 $P(x_i^T, x_j^T|x^0)$ accurately but to give a “functional” parametrization of $P(x^T|x^0)$

Alfredo Braunstein

Bayesian inference of cascades on networks

Politecnico di Torino
The problem and classical approaches

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 $P(x_i^T, x_j^T | x^0)$ accurately but to give a “functional” parametrization of $P(x^T | x^0)$
- [Note: to recover the MRF independence property one should fix full columns/trajectories $x_i^{0:T}$]
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. $x_i = SSSIRRRRR$ 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. $x_i = \text{SSSI}R\text{RRRR}$ 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. $x_i = \text{SSSI} \text{RRRR}$ 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 $P(s_{ij} = 0) = \lambda$
Parametrization of trajectories

- We will assume for simplicity $\mu = 1$. The case $\mu < 1$ is similar.
- Single site trajectories, e.g. $x_i = \text{SSS}1\text{RRRRR}$ 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 $P(s_{ij} = 0) = \lambda$
  1. Afterwards, all $t_i \neq 0$ can be computed *deterministically* with the following self-consistling equations

$$t_i = 1 + \min_{j \in \partial i} \{ t_j + s_{ji} \}$$
Parametrization of trajectories

- We will assume for simplicity $\mu = 1$. The case $\mu < 1$ is similar.
- Single site trajectories, e.g. $\mathbf{x}_i = SSS|IRRRRR$ 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 $P(s_{ij} = 0) = \lambda$

2. Afterwards, all $t_i \neq 0$ can be computed *deterministically* with the following self-consisting equations

$$t_i = 1 + \min_{j \in \partial_i} \{t_j + s_{ji}\}$$

- Key: stochastic parameters are *independent*
A static representation of SIR

A factorized distribution

\[ \mathcal{P}(t|x^0) = \sum_s \mathcal{P}(t|s, x^0) \mathcal{P}(s) \]
A static representation of SIR

A factorized distribution

\[ P(t|x^0) = \sum_s P(t|s, x^0) P(s) \]

Define

- \( \omega_{ij}(s_{ij}) = \lambda \delta(s_{ij}, 0) + (1 - \lambda) \delta(s_{ij}, \infty) \)
- \( \phi_i(t_i, t_{\partial i}, s_{\partial i}, x^0_i) = \delta(t_i, \delta(x^0_i; 1) (1 + \min_{j \in \partial i} \{ t_j + s_{ji} \})) \)
A static representation of SIR

A factorized distribution

\[ P(t|x^0) = \sum_s P(t|s, x^0) P(s) \]

Define

- \( \omega_{ij}(s_{ij}) = \lambda \delta(s_{ij}, 0) + (1 - \lambda) \delta(s_{ij}, \infty) \)
- \( \phi_i(t_i, t_{\partial i}, s_{\partial i}, x^0_i) = \delta(t_i, \delta(x^0_i; I)(1 + \min_{j \in \partial i} \{t_j + s_{ji}\})) \)

\[ Q = \frac{1}{Z} \prod_i \phi_i \prod_{i,j} \omega_{ij} \]
A static representation of SIR

A factorized distribution

\[
P(t|x^0) = \sum_s P(t|s,x^0)P(s)
\]

Define

\[\omega_{ij}(s_{ij}) = \lambda \delta(s_{ij}, 0) + (1 - \lambda) \delta(s_{ij}, \infty)\]

\[\phi_i(t_i, t_{\partial i}, s_{\partial i}, x^0_i) = \delta(t_i, \delta(x^0_i; I) (1 + \min_{j \in \partial i} \{ t_j + s_{ji} \}))\]

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

Then \[P(t|x^0) = \sum_s Q(t, s, x^0)\]
Adding priors

- \(x^T\) depends deterministically on \(t\): 
  \[ P(x^T | t) = \prod_i \xi_i(t_i, x_i^T) \]

where \(\xi_i(t_i, x_i^T)\) is the indicator function of

\[
\left( x_i^T = S, t_i > T \right) \lor \left( x_i^T = I, t_i = T \right) \lor \left( x_i^T = R, t_i < T \right)
\]
A static representation of SIR

Adding priors

- $x^T$ depends **deterministically** on $t$: $P(x^T|t) = \prod_i \xi_i(t_i, x^T_i)$ where $\xi_i(t_i, x^T_i)$ is the indicator function of

$$
(x^T_i = S, t_i > T) \lor (x^T_i = I, t_i = T) \lor (x^T_i = R, t_i < T)
$$

- $x^0$ have a prior concentrated on single-seed initial conditions: $P(x^0) = \prod_i \gamma_i(x^0_i)$ with $\gamma_i(x^0_i = I)$ very small.
Adding priors

- $x^T$ depends deterministically on $t$: $P(x^T | t) = \prod_i \xi_i(t_i, x_i^T)$ where
  $\xi_i(t_i, x_i^T)$ is the indicator function of
  
  $\left( x_i^T = S, t_i > T \right) \lor \left( x_i^T = I, t_i = T \right) \lor \left( x_i^T = R, t_i < T \right)$

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

- Finally, we can write the posterior distribution
  
  $P(x^0 | x^T) \propto \sum_t P(x^T | t) P(t | x^0) P(x^0)$ as

  $$P(x^0 | x^T) \propto \sum_t \sum_s \prod_{ij} \phi_{ij} \prod_i \phi_i \xi_i \gamma_i$$ (1)
Belief Propagation

\[ P(x^0|x^T) \propto \sum_t \sum_s \left[ \prod_{ij} \phi_{ij} \prod_i \phi_i \xi_i \gamma_i \right] = \sum_t \sum_s Q(x^0, t, s) \]

Single-instance RS cavity equations / Belief Propagation

- Fixed-point equation \( m = F_{BP}(m) \) for a vector \( m \) (called cavity marginals or messages) that is solved by iteration.
  - On a fixed point (approximate) marginals \( P(t_i|x^T) \) or \( P(x^0_i|x^T) \) 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

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

proximity [Isella et al.’10]  
sexual [Rocha et al’10]
Inferring $\lambda$ and $\mu$

Inference of parameters

- The likelihood of $\lambda, \mu$ can be computed as:

$$P(x^T | \lambda, \mu) = \sum_{t,g,x^0} P(x^T | t, g) P(t, g | x^0, \lambda, \mu) P(x^0)$$
Inferring $\lambda$ and $\mu$

Inference of parameters

- The likelihood of $\lambda, \mu$ can be computed as:

$$P(x^T | \lambda, \mu) = \sum_{t,g,x^0} P(x^T | t, g) P(t, g | x^0, \lambda, \mu) P(x^0) = Z$$
Inference of parameters

- The likelihood of $\lambda, \mu$ can be computed as:

\[
P(x^T|\lambda, \mu) = \sum_{t, g, x^0} P(x^T | t, g) P(t, g | x^0, \lambda, \mu) P(x^0) = Z \sim Z_{\text{Bethe}}
\]
Inferring $\lambda$ and $\mu$

**Inference of parameters**

- The likelihood of $\lambda, \mu$ can be computed as:
  \[
P(X^T | \lambda, \mu) = \sum_{t,g,x^0} P(X^T | t, g) P(t, g | x^0, \lambda, \mu) P(x^0) = Z \sim Z_{\text{Bethe}}
  \]
Interleaved BP+GA

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

\[
\frac{\partial}{\partial \lambda} [f (\mathbf{m}, \lambda)] = \nabla_{\mathbf{m}} f \cdot \frac{\partial \mathbf{m}}{\partial \lambda} + \frac{\partial f}{\partial \lambda}
\]
Interleaved BP+GA

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

$$\frac{\partial}{\partial \lambda} [f(m, \lambda)] = \nabla_m f \cdot \frac{\partial m}{\partial \lambda} + \frac{\partial f}{\partial \lambda} = \frac{\partial f}{\partial \lambda}$$

because $\nabla_m f \equiv 0$ on a FP of BP, as the BP solution is a variational critical point of $f$. 
Inferring $\lambda$ and $\mu$

Interleaved BP+GA

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

$$\frac{\partial}{\partial \lambda} [f (m, \lambda)] = \nabla_m f \cdot \frac{\partial m}{\partial \lambda} + \frac{\partial f}{\partial \lambda} = \frac{\partial f}{\partial \lambda}$$

because $\nabla_m f \equiv 0$ on a FP of BP, as the BP solution is a variational critical point of $f$.

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

the computation of an observable
Inferring $\lambda$ and $\mu$

**Interleaved BP+GA**

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

  $$\frac{\partial}{\partial \lambda} [f(m, \lambda)] = \nabla_m f \cdot \frac{\partial m}{\partial \lambda} + \frac{\partial f}{\partial \lambda} = \frac{\partial f}{\partial \lambda}$$

  because $\nabla_m f \equiv 0$ on a FP of BP, as the BP solution is a variational critical point of $f$.

- Now $\frac{\partial f}{\partial \lambda} (m, \lambda) = -\frac{1}{Z} \frac{\partial}{\partial \lambda} \left\{ \sum_{t,s} e^{\sum_i \log \psi_i + \sum_{\langle ij \rangle} \log \phi_{ij}} \right\} = -\sum_{t,s} \sum_{\langle ij \rangle} \left\{ \frac{\partial}{\partial \lambda} \log \phi_{ij} \right\} \left[ \frac{1}{Z} e^{\sum_i \log \psi_i + \sum_{\langle ij \rangle} \log \phi_{ij}} \right] = -\sum_{\langle ij \rangle} \left\langle \frac{\partial}{\partial \lambda} \log \phi_{ij} \right\rangle$, 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
Inferring $\lambda$ and $\mu$

**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 $\lambda$
Inferring $\lambda$ and $\mu$

Inference of network topology

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

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

- ROC area with \(N(N - 1)/2\) points using sorted inferred values \(\lambda_{ij}\)
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**

Thank you!