Supplementary Material for Paper: Discriminative Learning of Infection Models

1 Independent Cascade Model

The independent cascade (IC) model dynamics is equivalent to the following. First, sample each edge $e \in E$ to be *active* independently with probability θ_e. Then, the infected nodes are those reachable in the active network from some node in the seed s.

The integral in the kernel entries can be simplified (for simplicity we assume $\Theta = [0,1]^{|E|}$):

$$
\begin{aligned}
K(x,x') &= \int_\Theta f(s;\theta)f(s';\theta)\,\mathrm{d}\theta = \int_\Theta \mathbb{E}_\theta[y|s]\mathbb{E}_\theta[y|s']\,\mathrm{d}\theta \\
&= \int_\Theta \left(\sum_{A\subseteq E} P_\theta(A)r(A,s)\right)\left(\sum_{A'\subseteq E} P_\theta(A')r(A',s')\right)\mathrm{d}\theta \\
&= \sum_{A,A'\subseteq E} r(A,s)r(A',s')\int_\Theta P_\theta(A)P_\theta(A')\,\mathrm{d}\theta
\end{aligned}
$$

With a unifrom prior over θ, marginalizing over θ gives us a joint distribution Q over A, A':

$$
Q(A,A') = \int_\Theta P(A,A',\theta)\,\mathrm{d}\theta = \int_\Theta P(A|\theta)P(A'|\theta)P(\theta)\,\mathrm{d}\theta = \frac{1}{|\Theta|}\int_\Theta P_\theta(A)P_\theta(A')\,\mathrm{d}\theta
$$

Now we can write:

$$
K(x,x') \propto \sum_{A,A'\subseteq E} r(A,s)r(A',s')q(A,A') = \mathbb{E}_q[r(A,s)r(A',s')]
$$

When each θ_e is integrated over $[a,b]$, using edge independence we get:

$$
\begin{aligned}
Q(A,A') &= \frac{1}{|\Theta|}\int_\Theta P_\theta(A)P_\theta(A')\,\mathrm{d}\theta \\
&= \frac{1}{|\Theta|}\int_\Theta \prod_{e\in A}\theta_e \prod_{\bar{e}\notin A}(1-\theta_{\bar{e}}) \prod_{e'\in A'}\theta_{e'} \prod_{\bar{e}'\notin A'}(1-\theta_{\bar{e}'})\,\mathrm{d}\theta \\
&= \frac{1}{(b-a)^{|E|}}\int_\Theta \prod_{e\in A\cap A'}\theta_e^2 \prod_{\tilde{e}\in A\setminus A'}\theta_{\tilde{e}}(1-\theta_{\tilde{e}}) \prod_{\hat{e}\in A'\setminus A}\theta_{\hat{e}}(1-\theta_{\hat{e}}) \prod_{\bar{e}\in\overline{A\cup A'}}(1-\theta_{\bar{e}})^2\,\mathrm{d}\theta \\
&= \left(\frac{1}{b-a}\int_a^b u^2 du\right)^{|A\cap A'|}\left(\frac{1}{b-a}\int_a^b u(1-u)du\right)^{|A\setminus A'|+|A'\setminus A|}\left(\frac{1}{b-a}\int_a^b (1-u)^2 du\right)^{|\overline{A\cup A'}|} \\
&= q_\cap^{|A\cap A'|}q_\oplus^{|A\setminus A'|}q_\oplus^{|A'\setminus A|}q_{\bar{\cup}}^{|\overline{A\cup A'}|}
\end{aligned}
$$

where $q_\cap = \frac{b^3-a^3}{3(b-a)}$, $q_\oplus = \frac{b^2-a^2}{2(b-a)} - \frac{b^3-a^3}{3(b-a)}$, and $q_{\bar{\cup}} = \frac{b^3-a^3}{3(b-a)} - \frac{b^2-a^2}{b-a} + 1$. This suggest that under q, each edge is independently drawn to be in both A and A' with probability $q_\cap$, in either A or A' (but not in both) with probability $q_{\oplus}$, and in neither with probability $q_{\bar{\cup}}$. Hence, the computational dificulty of computing the kernel entries reduces from estimating influence over IC model parametrization given by all (or a large smaple of) θ-s to simply estimating 'co-influence' under a single 'co-IC' parametrization given by q.

Grouped Cascade Model In this generalization of the IC model, each node is a-priori assigned to one of K groups. As in the IC model, when a node i is infected, he has a single attempt to infect each of his non-infected neighbors j. In the grouped cascade model (GC), neighbors in the same group are "tied" - i independently infects *all* nodes in a group $k \in [K]$ with probability θ_{ik}, and *none* with probability $1 - \theta_{ik}$. When for every node i each neighbor j is in a different group (for instance when $K = n$), this model is equivalent to the IC model. Computing the kernel matrix for the GC model is similar to the IC model - instead of having a product over edges (i, j), our product will be over node-group pairs (i, k). To sample active edge sets A, A' under this model, for each i and $k \in [K]$ we sample all edges $\{(i, j) \in E : g(j) = k\}$ to be in both A and A', only one of them, or neither of them, according to Q^{IC}_{ik}.

2 Linear Threshold Model

In the linear threshold (LT) model, each neighbor j of node i has weight θ_{ji}, with $\sum_j \theta_{ji} \le 1$. In each instance, each node i first samples a threshold $\tau_i \sim U[0, 1]$. Then, a node i is infected if the sum of weights of his infected neighbors is higher than his threshold.
An equivalent process is as follows: First, for each node i, at most one incoming edge (j, i) becomes active with probability θ_{ji} (and no edge is active with probability $1 - \sum_j \theta_{ji}$). Then, the set of infected nodes are those for which there exists an active path from some seed node.

The kernel entries can be simplified. Denote by S^d the d-dimensional simplex, $S_i = S^{|N(i)|}$, and $\Theta = S_1 \times \cdots \times S_n$. As in the independent cascade kernel, the challenge is computing the integral over the product of probabilities. First, consider a joint distribution over (A, A', θ) with a uniform prior $P(\theta) = \frac{1}{|\Theta|} = 1/\prod_i |S_i|$, and A, A' are independent given θ. By marginalizing out θ we can write:

$$
\begin{aligned}
P(A, A') &= \int_\Theta P(A, A', \Theta)\, \mathrm{d}\theta = \int_\Theta P(A|\theta) P(A'|\theta) P(\theta)\, \mathrm{d}\theta \\
&= \prod_i \frac{1}{|S_i|} \int_{S_i} P_\theta(A_{\cdot i}) P_\theta(A'_{\cdot i})\, \mathrm{d}\theta_{\cdot i} = \prod_i P(A_{\cdot i}, A'_{\cdot i})\, \mathrm{d}\theta
\end{aligned}
$$

Recall that at most one entry in $A_{\cdot i}$ is on - denote it by $A(i) \in [N(i)] \cup \phi$ (and $A'(i)$ for A'). When $A(i) \neq A'(i)$, we have:

$$
\frac{1}{|S_i|} \int_{S_i} P_\theta(A_{\cdot i}) P_\theta(A'_{\cdot i})\, \mathrm{d}\theta_{\cdot i} = d! \int_{S^d} u_1 \cdot u_2\, \mathrm{d}u_1 \ldots \mathrm{d}u_d = \frac{d!}{(d+2)!} = \frac{1}{(d+1)(d+2)}
$$

and when $A(i) = A'(i)$, we have:

$$
\frac{1}{|S_i|} \int_{S_i} P_\theta(A_{\cdot i}) P_\theta(A'_{\cdot i})\, \mathrm{d}\theta_{\cdot i} = d! \int_{S^d} (u_1)^2\, \mathrm{d}u_1 \ldots \mathrm{d}u_d = 2\frac{d!}{(d+2)!} = \frac{2}{(d+1)(d+2)}
$$

since[1]:

$$
\int_{S^d} u_1^{\ell_1} \cdot \ldots \cdot u_d^{\ell_d} \cdot (1 - \sum_{k=1}^{d} u_k)^{\ell_0} = \frac{\prod_{k=0}^{d} \ell_k!}{\left(\sum_{k=0}^{d} \ell_k + d\right)!}
$$

[1] http://math.stackexchange.com/questions/207073/definite-integral-over-a-simplex

Once again we attained a distribution from which we can sample co-influences, this time under the linear threshold model: for each node i, we draw a joint edge (or no edges) with probability $\frac{2}{(d+1)(d+2)}$, and different edges (or no edges) with probability $\frac{1}{(d+1)(d+2)}$.

3 Branching Pocesses

In a branching process, a tree is generated by recursively sampling the number of children a node in the tree has from some fixed distribution. We use multi-type branching processes, where a node can either be *active* (and can have children) or *passive* (and cannot have children). The expected number of offspring μ_ℓ at depth ℓ is given by recursion:

$$\mu_0 = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad M = \begin{pmatrix} \mu_a & \mu_p \\ 0 & 0 \end{pmatrix}, \quad \mu_\ell = \mu_{\ell-1} \cdot M = \mu_0 \cdot M^\ell \tag{1}$$

Where μ_a, μ_p are the expected number of active and passive children of a given node. A node i receives $1/2^\ell$ credit for each offspring at distance ℓ (with maximal depth δ). The expected credit of i is therefore:

$$\text{credit}(i) = c(x^i) = \sum_{\ell=1}^{\delta} \frac{1}{2^\ell}(\mu_a^\ell + \mu_a^{\ell-1}\mu_p) \tag{2}$$

When the number of children are drawn from the geometric distribution with parameters θ, we get $\mu = \frac{1}{\theta} - 1$. In our model, θ is set to be:

$$\theta(x; w,) = \frac{1}{1 + e^{-\langle w,x \rangle + b}} \quad \Rightarrow \quad \mu = e^{-\langle w,x \rangle + b} \tag{3}$$

where x are a node's features and w, b are weights. We present a closed form solution for the kernel entries. Define $\mathbf{w} = (w_a, b_a, w_p, b_p)$, then:

$$\begin{aligned} K(x,x') &= \int c(x;w)c(x';w)\,\mathrm{d}\mathbf{w} = \int \left(\sum_{\ell=1}^{\delta} \frac{1}{2^\ell}(\mu_a^\ell + \mu_a^{\ell-1}\mu_p) \right) \left(\sum_{\ell=1}^{\delta} \frac{1}{2^\ell}(\mu_a'^{\ell'} + \mu_a'^{\ell'-1}\mu_p') \right) \mathrm{d}\mathbf{w} \\ &= \sum_{\ell,\ell'=1}^{\delta} \frac{1}{2^{\ell+\ell'}} \int (\mu_a^\ell + \mu_a^{\ell-1}\mu_p)(\mu_a'^{\ell'} + \mu_a'^{\ell'-1}\mu_p')\,\mathrm{d}\mathbf{w} \\ &= \sum_{\ell,\ell'=1}^{\delta} \frac{1}{2^{\ell+\ell'}} \left(\underbrace{\int \mu_a^\ell \mu_a'^{\ell'}\,\mathrm{d}\mathbf{w}}_{\text{(I)}} + \underbrace{\int \mu_a^\ell \mu_a'^{\ell'-1}\mu_p'\,\mathrm{d}\mathbf{w}}_{\text{(II)}} + \underbrace{\int \mu_a'^{\ell'} \mu_a^{\ell-1}\mu_p\,\mathrm{d}\mathbf{w}}_{\text{(II)'}} + \underbrace{\int \mu_a^{\ell-1}\mu_p \mu_a'^{\ell'-1}\mu_p'\,\mathrm{d}\mathbf{w}}_{\text{(III)}} \right) \end{aligned}$$

with:

$$\begin{aligned} \text{(I)} &= \int \left(e^{-\langle w_a,x \rangle + b_a}\right)^\ell \left(e^{-\langle w_a,x' \rangle + b_a}\right)^{\ell'} \mathrm{d}\mathbf{w} \\ &= \int \mathrm{d}w_p \int \mathrm{d}b_p \int e^{-\langle w_a, \ell x + \ell' x' \rangle + b_a(\ell+\ell')}\,\mathrm{d}w_a\,\mathrm{d}b_a \\ &= V_{w_p} V_{b_p} \int e^{b_a(\ell+\ell')}\,\mathrm{d}b_a \prod_k \int e^{-(w_a)_k \overbrace{(\ell(x_a)_k + \ell'(x_a')_k)}^{\alpha_k^1}}\,\mathrm{d}(w_a)_k \\ &= V_{w_p} V_{b_p} \frac{1}{\ell+\ell'} \left[e^{b_a(\ell+\ell')} \right]_{b_a^{\min}}^{b_a^{\max}} V_{w_a}^{|\{k:\alpha_k^1=0\}|} s \prod_{k:\alpha_k^1 \neq 0} \frac{1}{\alpha_k^1} \left[e^{-(w_a)_k \alpha_k^1} \right]_{(w_a)_k^{\max}}^{(w_a)_k^{\min}} \end{aligned} \tag{$*$}$$

$$
\begin{aligned}
\text{(II)} &= \int \left(e^{-\langle\, w_a, x\,\rangle + b_a}\right)^{\ell} \left(e^{-\langle\, w_a, x'\,\rangle + b_a}\right)^{\ell'-1} e^{-\langle\, w_p, x'\,\rangle + b_p} \,\mathrm{d}\mathbf{w} \\
&= \int e^{b_p}\,\mathrm{d}b_p \int e^{b_a(\ell+\ell'-1)}\,\mathrm{d}b_a \prod_k \int e^{-(w_p)_k x'_k}\,\mathrm{d}(w_p)_k \int e^{-(w_a)_k \overbrace{(\ell x_k + (\ell'-1)x'_k)}^{\alpha_k^2}}\,\mathrm{d}(w_a)_k \\
&= \left[e^{b_p}\right]_{b_p^{\min}}^{b_p^{\max}} \frac{1}{\ell+\ell'-1} \left[e^{b_a(\ell+\ell'-1)}\right]_{b_a^{\min}}^{b_a^{\max}} \cdots \qquad (*) \\
&\qquad V_{w_p}^{|\{k:x'_k=0\}|} \prod_{k:x'_K\neq 0} \frac{1}{x'_k} \left[e^{-(w_p)_k x'_k}\right]_{(w_p)_k^{\max}}^{(w_p)_k^{\min}} V_{w_a}^{|\{k:\alpha_k^2=0\}|} \prod_{k:\alpha_k^2\neq 0} \frac{1}{\alpha_k^2} \left[e^{-(w_a)_k \alpha_k^2}\right]_{(w_a)_k^{\max}}^{(w_a)_k^{\min}}
\end{aligned}
$$

$$
\begin{aligned}
\text{(III)} &= \int \left(e^{-\langle\, w_a, x\,\rangle + b_a}\right)^{\ell-1} e^{-\langle\, w_p, x\,\rangle + b_p} \left(e^{-\langle\, w_a, x'\,\rangle + b_a}\right)^{\ell'-1} e^{-\langle\, w_p, x'\,\rangle + b_p} \\
&= \int e^{2b_p}\,\mathrm{d}b_p \int e^{(\ell+\ell'-2)b_a}\,\mathrm{d}b_a \prod_k \int e^{-(w_p)_k (x_k + x'_k)}\,\mathrm{d}(w_p)_k \int e^{-(w_a)_k \overbrace{((\ell-1)x_k + (\ell'-1)x'_k)}^{\alpha_k^3}}\,\mathrm{d}(w_a)_k \\
&= \frac{1}{2}\left[e^{2b_p}\right]_{b_p^{\min}}^{b_p^{\max}} \frac{1}{\ell+\ell'-2} \left[e^{b_a(\ell+\ell'-2)}\right]_{b_a^{\min}}^{b_a^{\max}} \cdots \qquad (*) \\
&\qquad V_{w_p}^{|\{k:x_k+x'_k=0\}|} \prod_{k:x_k+x'_k\neq 0} \frac{1}{(x_k+x'_k)} \left[e^{-(w_p)_k (x_k+x'_k)}\right]_{(w_p)_k^{\max}}^{(w_p)_k^{\min}} \cdots \\
&\qquad V_{w_a}^{|\{k:\alpha_k^3=0\}|} \prod_{k:\alpha_k^3\neq 0} \frac{1}{\alpha_k^3} \left[e^{-(w_a)_k \alpha_k^3}\right]_{(w_a)_k^{\max}}^{(w_a)_k^{\min}}
\end{aligned}
$$

$(*)$ when the denominator is 0, $[e^{-x}]_{x^{\min}}^{x^{\max}}$ is replaced with $x^{\max} - x^{\min}$