Assignment 02

Theory of social computing | asj 170430.

Q1

There are two phases for TIM algorithm.

1) parameter estimation
2) Node selection phase.

A_0 — seed set

A_K — infected node set at stage K.

$$\text{Prob}\begin{bmatrix} u \text{ not infected at } K \\ \text{but infected at } K+1 \end{bmatrix} = \frac{\sum_{w \in A_K | A_{K-1}} b_{wu}}{1 - \sum_{w \in A_{K-1}} b_{wu}}$$

i.e. $\text{Prob}\begin{bmatrix} u \text{ not infected at } K \\ \text{but infected at } K+1 \end{bmatrix}$

$$= \frac{\Pr[\exists w \text{ that } w \text{ makes } u \text{ active}]}{\Pr[\forall w \text{ that din't make } u \text{ active}]}$$

∴ 1) node sampled uniformly at random from G

2) a random graph obtained from removing an edge with $1 - \begin{bmatrix} \text{Prob } u \text{ not infe. at } K \\ \text{but infected at } K+1 \end{bmatrix}$.

Q2

a randomized bfs is used for generation of each RR.

we randomly choose edge e of v and add node u at queue.

The problem with randomized dfs

1) it is not optimal in finding shortest distance

2) the social graph is not DAG (i.e. directed acyclic graph and hence dfs can get stuck in infinite loop.

Q4 Proof of lemma 3.

let ρ be the probability that S overlaps with random RR set then $\theta . F_R(S)$ can be regarded as the sum of θ i.i.d bernoulli variable with a mean ρ.

$$\rho = E[F_R(S)] = E[I(S)]/n$$

then we have

$$\Pr\left[\left| n \cdot F_R(S) - E[I(S)] \right| \geq \delta/2 \cdot OPT \right]$$

$$= \Pr\left[\left| \theta \cdot F_R(S) - \rho\theta \right| \geq \frac{\varepsilon\theta}{2n} \cdot OPT \right]$$

$$= \Pr\left[\left| \theta \cdot F_R(S) - \rho\theta \right| \geq \frac{\varepsilon \cdot OPT}{2n\rho} \cdot \rho\theta \right] \quad -(1).$$

let $\delta = \varepsilon \cdot OPT / 2n\rho$. by chernoff bounds equation 2 and fact that $\rho = E[I(S)]/n \leq OPT/n$ we have

$$\text{RHS of equ (1)} < 2 \cdot \exp\left(-\frac{\delta^2}{2+\delta} \cdot \rho\theta \right).$$

$$< 2 \cdot \exp\left(-\frac{\varepsilon^2 OPT^2}{8n^2\rho + 2\varepsilon n \cdot OPT} \cdot \theta \right)$$

$$< 2\exp\left(-\frac{\varepsilon^2 \cdot OPT^2}{8n\, OPT + 2\varepsilon n \cdot OPT} \cdot \theta \right)$$

$$< \frac{1}{\binom{n}{k} \cdot n^2}$$

Hence lemma is proved.

let $\delta(2^{-i}-\mu)/\mu$ by chernoff bounds

$$\Pr\left[\frac{\delta_i}{c_i} > 2^{-i}\right] \leq \exp\left(\frac{-\delta^2}{2+\delta} \cdot c_i \cdot \mu\right)$$

$$= \exp\left(-c_i \times (2^{-i}-\mu)^2 / (2^{-i}+\mu)\right)$$

$$< \exp\left(-c_i \cdot 2^{-i-1}/3\right)$$

$$= \frac{1}{n^{\lambda} \cdot \log_2 n}$$

Q5 advantages.

1) it provides the same worst case quarantees as the state of art but effects significantly improved accuracy and empirically efficiency

2) the core is based on martingales a classic statistical tool which not only provides accurate result but with

small computation. and also supports large set of diffusion models.

Q3.

$\hat{p} = \frac{X}{n}$ ← binomial random variable.
← total # of individuals in sample.

discrete random variable can take $0, 1/n, 2/n, \ldots, 1$ values.

binomial random var.

mean $= np$. & $Var(X) = n.p.(1-p)$.

$E(\hat{p}) = E\left(\frac{X}{n}\right)$

$= 1/n \; E(X) = \frac{1}{n} \cdot n \cdot p = p$

$\hat{p}$ ← unbiased approx. of p.

$\hat{\sigma}_p = Var(\hat{p}) = Var(X/n)$

$= 1/n^2 \cdot Var(X)$

$= 1/n^2 \cdot n \cdot p \cdot (1-p)$.

$$= \frac{p\cdot(1-p)}{n}$$

$$\therefore \sigma_{\hat{p}} = \sqrt{\frac{1}{n}\cdot p(1-p)}$$

sampling distribution of $\hat{p}$, is approximately normal if

$np \geq 5$ & $n(1-p) \geq 5$