12/3/2022

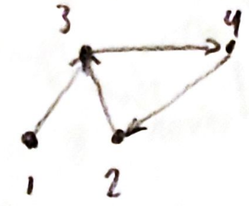
Maybe I can use the graphon $W(u,v) \equiv p \in [0,1]$ to understand what n and $|A|$ should be.

- We need $|A|$ large enough that, ~~the~~ for all network-wide ~~properties~~ ~~—~~ measures we want to test, the empirical distribution of those measures is "close enough" ~~to~~ to the graphon distribution.
- Similarly, we need n large enough that for all nodewise measures ~~of~~ we want to test, the empirical distribution of those measures is close enough to the ~~empirical~~ distribution predicted by the graphon (note that we'll have to choose a clever labeling of the nodes).
- Also, am I giving the same treatment to each message?

What I mean is:

→

Take the following example

A = 

$L = 2$

$x_0 = \begin{bmatrix}0\\0\\0\\0\end{bmatrix}, b_0 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix};$

$x_1 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix}, b_1 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix} \Rightarrow Ax_1 = \begin{bmatrix}0\\0\\2\\0\end{bmatrix}, Ax_1 + b_1 = \begin{bmatrix}1\\1\\2\\0\end{bmatrix}$

just node count

so $x_2 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix}$, and we know the lifetime of message $m_{1,\ell} \geq 2$

but $x_0 = \begin{bmatrix}0\\0\\0\\0\end{bmatrix}, b_0 = \begin{bmatrix}1\\0\\1\\0\end{bmatrix} \Rightarrow x_1 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}, b_1 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix}; x_2 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, Ax_2 = \begin{bmatrix}0\\1\\2\\1\end{bmatrix}, b_2 = \begin{bmatrix}0\\1\\1\\0\end{bmatrix};$

$x_3 = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$ so the lifetime of message $m_{1,1} \geq 3$

So it seems like injecting a message at node 1 via $\begin{bmatrix}1\\1\\0\\0\end{bmatrix}$ is a different treatment than doing so via $\begin{bmatrix}1\\0\\1\\0\end{bmatrix}$.

I think it will help me to see what the assumptions of the Kaplan-Meier curve are.

No assumptions listed in book.

- I guess what I'm saying is that the extinction time of a message depends on the locations where the other messages are injected.
- ~~My hope is that it would just "average out"~~

- ☐ Finish week 2
- ★ ☐ Finish code
- ☐ 1/2 week 3 (more if feel)
- ☐ Excel practice
- ☐ Read those 2 papers

$S(t) = \mathbb{P}(T>t)$ $\qquad$ $T_{i,\alpha}$ = survival time of the α-th message injected at node i

$$\mathbb{P}(T_{i,1} > t) = \mathbb{P}(T_{i,1} > t \mid b_0, x_{-1})$$

12/3/2022

$t(F,W) = \int$
$_{[0]}$

Let W be a graphon and F be a motif with vertex set V' and edge set E'. Then

$$t(F,W) = \int_{[0,1]^{|V'|}} \prod_{(i,j)\in E'} W(u_i,u_j) \prod_{i\in V'} du_i$$

is the probability of drawing motif F from graphon W

Try just an edge.

$\int_{[0,1]^2} p \, du dv = p \int_{[0,1]^2} du dv = p$ which is also the expected edge density

Try a triangle

$\int_{[0,1]^3} p^3 \, du dv dw = p^3$

Try a 2-path

$\int_{[0,1]^3} p^2 \, du dv dw = p^2$

Hmm. harder than I thought.

So these are expected subgraph densities in the $n \to \infty$ limit. What about standard deviations? Presumably, the $SD \to 0$ as $n \to \infty$ for all subgraph densities due to convergence, so the question becomes: what is the rate of convergence?

A graph parameter f is estimable if $\forall \varepsilon>0\ \exists k\in\mathbb{N}$ s.t. if G is a graph with at least k nodes and we select a random k-set of nodes from G, then from the induced subgraph $G[X]$ we can estimate $g(G[X])$ of f s.t.

$$\mathbb{P}(|f(G)-g(G[X])|>\varepsilon)<\varepsilon$$

Graph parameters of the form $t(F,\cdot)$ are estimable.

What is ~~my~~ problem? I want an ensemble that has
1) A good spread in the distribution of graph parameters
2) Enough nodes to be interesting

What are some things we want in our global feature set?

- Number of nodes n
- Edge density
- Different probabilities p

Correlated

- Motif densities
- # connected components
- Characteristic path length
- Normalized rank

What about local feature set?

- In degree
- Out degree
- Centrality measures
- Motif membership?

Remember, small expriments first, you don't want to waste tons of time.

- [ ] Try purely local, purely global, and mixed local-global linear models.
- [ ] Be sure to visualize degree distributions
- [ ] Charge ahead despite ~~with~~ the concern that the treatments aren't the same.
- [ ] Do fixed load expriments and random load experiments (former to be comparable with Hao-Graham paper. latter due to ergodicity being trivial to prove)