May 2, 2019

Discrete choice & network growth

Last time:

logit model

$$U_{kj} = V_{kj} + \varepsilon_{kj}$$

$j \in$ choice set C_k

ε_{kj} i.i.d. Gumbel

choose $\arg\max_j U_{kj}$

$$\text{Prob}(\text{choose } j) = \frac{\exp(V_j)}{\sum_{\ell \in C_k} \exp(V_\ell)}$$

functional form for V

Today: $V_{kj} = \theta^T x_j$

x_j features of item j

Network growth

So far, assumed graph was fixed (static)

Graph $G \to$ clustering,
centrality,
learning

Useful to think about network assembly (graph growth over time)

Data:

$(u_1, v_1, t_1), \ldots, (u_N, v_N, t_N)$

Example: (generalized) preferential attachment u chooses to link to node v proportional to the degree of node v

Prob(u chooses v)

$$\left.= \frac{d_v^{\alpha}}{\sum_{(u,w)\in E} d_w^{\alpha}}\right\} \text{choice set depends on } u$$

Can think of this as logit

$$U_{u\to v} = \underset{\substack{\uparrow \\ \theta}}{\alpha} \log(\underset{\substack{\uparrow \\ x_v}}{d_v}) + \varepsilon$$

"Choosing to grow a graph"
(Overgoor, Benson, Ugander 2019)

Main idea: can get better estimates of growth mechanism by looking at temporal data

Note:
Utilities change as the graph grows (e.g. degree change)

features x_j really also depend on time

Other models

① Jackson-Rogers (2007)

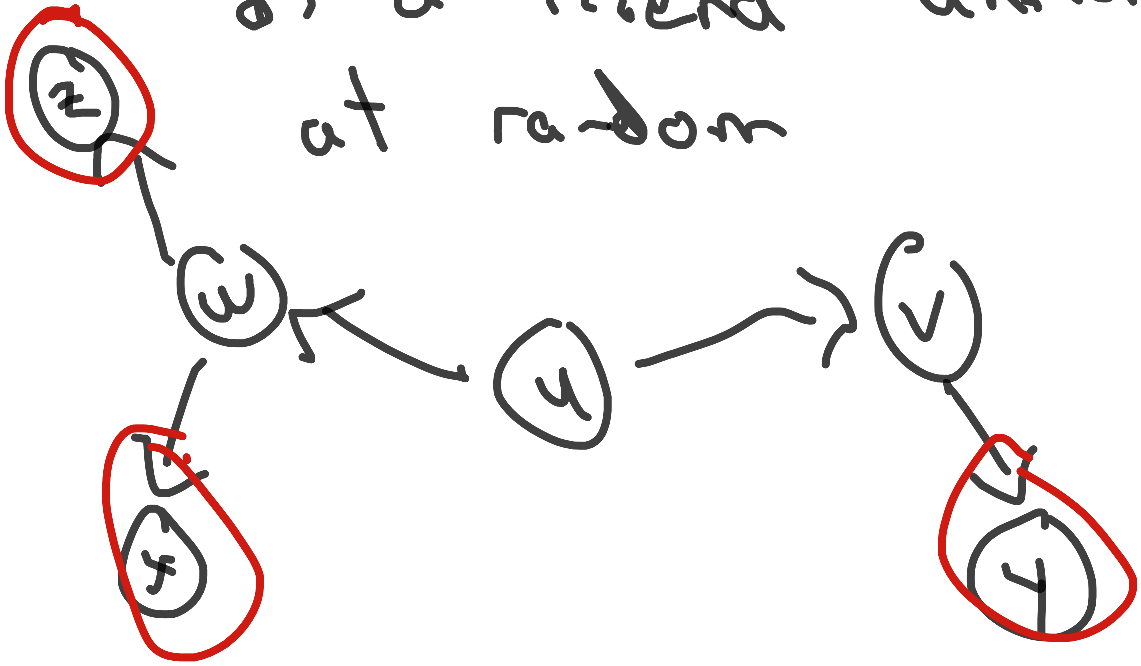
u is about to link

with prob. r:

u links to a randomly chosen node

with prob. 1-r:

u links to a "friend of a friend" uniformly at random



Base utility of any node a is $V_a = 0$

all that changes is choice set

w.p. r:

$$C = \{a \mid (u,a) \notin E\}$$

w.p. $1-r$:

$$C = \{a \mid (u,a) \notin E, \exists b \text{ s.t. } \underbrace{(u,b) \,\&\, (b,a)}_{\in E}\}$$

Example of a mixed logit with two components

② Fitness model

Each node given some latent fitness

$$f_v \sim D$$

Node u chooses proportional to fitness

$$\left.\Pr(u \text{ links to } v) = \frac{f_v}{\sum\limits_{(u,w)\notin E} f_w}\right\} \begin{array}{l}\text{Choice set}\\ = \text{nodes not}\\ \text{linked to } u\end{array}$$

can also give rise to heavy tailed degree distribution

$$f_v = \exp(V_v)$$

$$= \exp(\theta^T x_j)$$

$$x_j = \begin{array}{c c} & \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \end{pmatrix} \\ j & \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \end{array} \quad x_j(j) = 1$$

Can also use other covariates: age, political leaning, location, etc.

Likelihood ratio test

Example:

$(u_1, v_1, t_1), \ldots, (u_N, v_N, t_N)$

Model 1: preferential attachment

$$V_v = \theta \log(d_v)$$

Model 2: add age

$$V_v = \theta_1 \log(d_v) + \theta_2 \log(\text{how long } v \text{ has been in the network})$$

fit both models

likelihoods L_1, L_2

Null: Model 1

$$\underline{-2\log(L_1/L_2)} \sim \chi_1^2$$

Also can get things like confidence intervals on θ (under certain conditions)

Where to go from here?

Numerical linear algebra

CS 6210 (fall)
CS 6220 (spring)

More networks!

CS 6850 (spring)

Final projects due
4:30pm May 15