Learning Interpretable Feature Context Effects in Discrete Choice

Kiran Tomlinson
PhD Student, Cornell CS

research with Austin R. Benson

Code: bit.ly/lcl-code
Data: bit.ly/lcl-data
Slides: bit.ly/lcl-kdd-slides
Choices and context effects
Discrete choices are everywhere
“The fundamental problem of discrete choice”
“The fundamental problem of discrete choice”

choice set
“The fundamental problem of discrete choice”

choice set

choice

apple, lemon, grape, strawberry
“The fundamental problem of discrete choice”

choice set

choice

...
“The fundamental problem of discrete choice”
The classic model: *multinomial logit (MNL)*

(McFadden, *Frontiers in Econometrics* 1973)
The classic model: **multinomial logit (MNL)**

(McFadden, *Frontiers in Econometrics* 1973)

Assume *item* *i* has *utility* $u_i$

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$
The classic model: \textit{multinomial logit (MNL)}

(McFadden, Frontiers in Econometrics 1973)

Assume \textit{item} $i$ has \textit{utility} $u_i$

$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

<table>
<thead>
<tr>
<th>C</th>
<th>$u_i$</th>
<th>Pr($i \mid C$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>.24</td>
</tr>
<tr>
<td></td>
<td>-1</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.64</td>
</tr>
</tbody>
</table>
The classic model: **multinomial logit (MNL)**

(McFadden, *Frontiers in Econometrics* 1973)

Assume item $i$ has utility $u_i$

\[
Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}
\]

**Unique choice model satisfying independence of irrelevant alternatives (IIA):**

(Luce, *Individual Choice Behavior* 1959)

\[
\frac{Pr(i \mid C)}{Pr(j \mid C)} = \frac{Pr(i \mid C')}{Pr(j \mid C')}
\]
Problem for MNL: context effects
Problem for MNL: *context effects*

The choice set influences preferences.
Problem for MNL: *context effects*

The choice set influences preferences.

*Compromise*

(Simonson, 1989)
Problem for MNL: context effects

The choice set influences preferences.

Compromise

(Simonson, 1989)
Problem for MNL: context effects

The choice set influences preferences.

Compromise

(Simonson, 1989)
Problem for MNL: *context effects*

The choice set influences preferences.

*Compromise*
(Simonson, 1989)
Problem for MNL: *context effects*

The choice set influences preferences.

*Compromise*

(Simonson, 1989)
Problem for MNL: context effects

The choice set influences preferences.

Compromise  
(Simonson, 1989)

Similarity  
(Tversky, 1972)
Problem for MNL: **context effects**

The choice set influences preferences.

**Compromise**  
(Simonson, 1989)

- $10
- $15
- $20
- $25

**Similarity**  
(Tversky, 1972)

<table>
<thead>
<tr>
<th></th>
<th>Rep. 2,462,617</th>
<th>49.7%</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Perdue*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jon Ossoff</td>
<td>Dem. 2,374,519</td>
<td>47.9%</td>
</tr>
</tbody>
</table>

*(Headings may not align due to formatting)*
Problem for MNL: **context effects**

The choice set influences preferences.

**Compromise**  
(Simonson, 1989)

<table>
<thead>
<tr>
<th>Price</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$10</td>
<td></td>
</tr>
<tr>
<td>$15</td>
<td></td>
</tr>
<tr>
<td>$20</td>
<td></td>
</tr>
<tr>
<td>$25</td>
<td></td>
</tr>
</tbody>
</table>

**Similarity**  
(Tversky, 1972)

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Party</th>
<th>Votes</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Perdue*</td>
<td>Rep.</td>
<td>2,462,617</td>
<td>49.7%</td>
</tr>
<tr>
<td>Jon Ossoff</td>
<td>Dem.</td>
<td>2,374,519</td>
<td>47.9</td>
</tr>
<tr>
<td>Raphael Warnock</td>
<td>Dem.</td>
<td>1,617,035</td>
<td>32.9%</td>
</tr>
<tr>
<td>Kelly Loeffler*</td>
<td>Rep.</td>
<td>1,273,214</td>
<td>25.9</td>
</tr>
<tr>
<td>Doug Collins</td>
<td>Rep.</td>
<td>980,454</td>
<td>20.0</td>
</tr>
</tbody>
</table>
Problem for MNL: context effects

The choice set influences preferences.

**Compromise**
(Simonson, 1989)

- $10
- $15
- $20
- $25

**Similarity**
(Tversky, 1972)

- David Perdue*
  - Rep.
  - 2,462,617
  - 49.7%
- Jon Ossoff
  - Dem.
  - 2,374,519
  - 47.9%
- Raphael Warnock
  - Dem.
  - 1,617,035
  - 32.9%
- Kelly Loeffler*
  - Rep.
  - 1,273,214
  - 25.9%
- Doug Collins
  - Rep.
  - 980,454
  - 20.0%

**IIA violations:**

\[
\frac{Pr(i \mid C)}{Pr(j \mid C)} \neq \frac{Pr(i \mid C')}{Pr(j \mid C')}
\]
Natural context effect model: **CDM**

(Seshadri, Peysakhovich, & Ugander, ICML 2019)
Natural context effect model: **CDM**

(Seshadri, Peysakhovich, & Ugander, ICML 2019)

Item $j$ exerts *pull* $u_{ij}$ on item $i$, item utility is sum of pulls:

$$\Pr(i \mid C) = \frac{\exp \left( \sum_{k \in C \setminus i} u_{ik} \right)}{\sum_{j \in C} \exp \left( \sum_{k \in C \setminus i} u_{jk} \right)}$$
Natural context effect model: **CDM**

(Seshadri, Peysakhovich, & Ugander, ICML 2019)

Item $j$ exerts *pull* $u_{ij}$ on item $i$, item utility is sum of pulls:

$$\Pr(i \mid C) = \frac{\exp \left( \sum_{k \in C \setminus i} u_{ik} \right)}{\sum_{j \in C} \exp \left( \sum_{k \in C \setminus i} u_{jk} \right)}$$

Assumes no higher-order effects
(i.e., context effects decompose additively into effects of items)
Natural context effect model: CDM
(Seshadri, Peysakhovich, & Ugander, ICML 2019)

Item $j$ exerts pull $u_{ij}$ on item $i$, item utility is sum of pulls:

$$\Pr(i \mid C) = \frac{\exp \left( \sum_{k \in C \setminus i} u_{ik} \right)}{\sum_{j \in C} \exp \left( \sum_{k \in C \setminus i} u_{jk} \right)}$$

Assumes no higher-order effects (i.e., context effects decompose additively into effects of items)
Item features and the LCL
Choice models with *item features*
Choice models with *item features*

So far, models have per-item parameters
Choice models with *item features*

So far, models have per-item parameters

→ can’t generalize to new items not in training set
Choice models with *item features*

So far, models have per-item parameters

→ can’t generalize to new items not in training set
→ hard to learn utilities for rare items
Choice models with *item features*

So far, models have per-item parameters

- can’t generalize to new items not in training set
- hard to learn utilities for rare items
- too many parameters with many items
Choice models with *item features*

So far, models have per-item parameters

→ can’t generalize to new items not in training set
→ hard to learn utilities for rare items
→ too many parameters with many items

Use item features:

- **genre:** drama,
  - `in_top_10`: True,
  - `has_new_episodes`: True,
  - `producer`: Netflix

- **genre:** comedy,
  - `in_top_10`: False,
  - `has_new_episodes`: False,
  - `producer`: NBC

- **genre:** drama,
  - `in_top_10`: True,
  - `has_new_episodes`: False,
  - `producer`: Netflix

- **genre:** reality,
  - `in_top_10`: True,
  - `has_new_episodes`: False,
  - `producer`: Banijay
MNL with item features: *conditional logit*
MNL with item features: \textit{conditional logit}

\textit{Feature vector } \mathbf{x}_i \in \mathbb{R}^d \text{ for each item } i

\textit{Preference vector } \mathbf{\theta} \in \mathbb{R}^d
MNL with item features: *conditional logit*

Feature vector \( x_i \in \mathbb{R}^d \) for each item \( i \)

Preference vector \( \theta \in \mathbb{R}^d \)

MNL:

\[
\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}
\]
MNL with item features: \textit{conditional logit}

Feature vector $x_i \in \mathbb{R}^d$ for each item $i$
Preference vector $\theta \in \mathbb{R}^d$

MNL:
$$\Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

Conditional logit:
$$\Pr(i \mid C) = \frac{\exp(\theta^T x_i)}{\sum_{j \in C} \exp(\theta^T x_j)}$$
MNL with item features: *conditional logit*

*Feature vector* $x_i \in \mathbb{R}^d$ for each item $i$

*Preference vector* $\theta \in \mathbb{R}^d$

MNL:

$$Pr(i \mid C) = \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

Conditional logit:

$$Pr(i \mid C) = \frac{\exp(\theta^T x_i)}{\sum_{j \in C} \exp(\theta^T x_j)}$$

*Preference coefficient* $\theta_k$ is easy to interpret: importance of the $k^{th}$ feature
Incorporating *feature context effects* into conditional logit
Incorporating *feature context effects* into conditional logit

Conditional logit utility: \( u_i = \theta^T x_i \)
Incorporating *feature context effects* into conditional logit

Conditional logit utility: \( u_i = \theta^T x_i \)  \hspace{1cm}  \text{Contextual utility:}  \ u_{i,C} = [\theta + F(C)]^T x_i
Incorporating \textit{feature context effects} into conditional logit

Conditional logit utility: $u_i = \theta^T x_i$  \hspace{1cm} Contextual utility: $u_{i,C} = [\theta + F(C)]^T x_i$

Simplifying assumptions on $F(C)$:
Incorporating *feature context effects* into conditional logit

Conditional logit utility: \( u_i = \theta^T x_i \)

Contextual utility: \( u_{i,C} = [\theta + F(C)]^T x_i \)

Simplifying assumptions on \( F(C) \):

1. **Additivity**: \( F(C) \propto \sum_{j \in C} f(x_j) \) for some function \( f \)
Incorporating *feature context effects* into conditional logit

Conditional logit utility: \( u_i = \theta^T x_i \)  

Contextual utility: \( u_{i,C} = [\theta + F(C)]^T x_i \)

Simplifying assumptions on \( F(C) \):

1. *Additivity*: \( F(C) \propto \sum_{j \in C} f(x_j) \) for some function \( f \)

2. *Dilution*: \( F(C) = \frac{1}{|C|} \sum_{j \in C} f(x_j) \)
Incorporating *feature context effects* into conditional logit

Conditional logit utility: \( u_i = \theta^T x_i \)  
Contextual utility: \( u_{i,C} = [\theta + F(C)]^T x_i \)

Simplifying assumptions on \( F(C) \):

1. **Additivity**: \( F(C) \propto \sum_{j \in C} f(x_j) \) for some function \( f \)

2. **Dilution**: \( F(C) = \frac{1}{|C|} \sum_{j \in C} f(x_j) \)

3. **Linearity**: \( f(x_j) = Ax_j \) for some matrix \( A \in \mathbb{R}^{d \times d} \)
Incorporating feature context effects into conditional logit

Conditional logit utility: \( u_i = \theta^T x_i \) \quad \rightarrow \quad Contextual utility: \( u_{i,C} = [\theta + F(C)]^T x_i \)

Simplifying assumptions on \( F(C) \):

1. **Additivity**: \( F(C) \propto \sum_{j \in C} f(x_j) \) for some function \( f \)

2. **Dilution**: \( F(C) = \frac{1}{|C|} \sum_{j \in C} f(x_j) \)

3. **Linearity**: \( f(x_j) = Ax_j \) for some matrix \( A \in \mathbb{R}^{d \times d} \)

\[ \rightarrow u_{i,C} = (\theta + Ax_C)^T x_i \quad \quad (x_C = \frac{1}{|C|} \sum_{j \in C} x_j \text{ is the mean feature vector}) \]
The *Linear Context Logit (LCL)*
The **Linear Context Logit (LCL)**

\[
\Pr(i \mid C) = \frac{\exp([\theta + Ax_C]^T x_i)}{\sum_{j \in C} \exp([\theta + Ax_C]^T x_j)}
\]
The **Linear Context Logit (LCL)**

\[
Pr(i \mid C) = \frac{\exp([\theta + Ax_C]^T x_i)}{\sum_{j \in C} \exp([\theta + Ax_C]^T x_j)}
\]

→ convex negative log-likelihood
The **Linear Context Logit (LCL)**

\[
Pr(i \mid C) = \frac{\exp([\theta + Ax_C]^T x_i)}{\sum_{j \in C} \exp([\theta + Ax_C]^T x_j)}
\]

→ convex negative log-likelihood

→ \( \theta \): base preference coefficients
The **Linear Context Logit (LCL)**

\[
Pr(i \mid C) = \frac{\exp([\theta + Ax_C]^T x_i)}{\sum_{j \in C} \exp([\theta + Ax_C]^T x_j)}
\]

→ convex negative log-likelihood

→ \( \theta \): base preference coefficients

→ \( A_{pq} > 0 \): when \( q \) is *high* in the choice set, \( p \) is *more* preferred
The **Linear Context Logit (LCL)**

\[
Pr(i \mid C) = \frac{\exp([\theta + Ax_C]^T x_i)}{\sum_{j \in C}\exp([\theta + Ax_C]^T x_j)}
\]

→ convex negative log-likelihood

→ \( \theta \): base preference coefficients

→ \( A_{pq} > 0 \): when \( q \) is high in the choice set, \( p \) is more preferred

→ \( A_{pq} < 0 \): when \( q \) is high in the choice set, \( p \) is less preferred
LCL example: restaurant selection
LCL example: restaurant selection
LCL example: restaurant selection

item features:
- price
- service speed
- wine selection
LCL example: restaurant selection

item features:
- price
- service speed
- wine selection

$C_1$, $C_2$, $C_3$
LCL example: restaurant selection

item features:
- price
- service speed
- wine selection

mean choice set price

C_1  C_2  C_3
LCL example: restaurant selection

item features:
- price
- service speed
- wine selection

mean choice set price
LCL example: restaurant selection

Item features:
- price
- service speed
- wine selection

Mean choice set price

Importance of:
- speed
- wine selection

Icons by icons8.com
LCL identifiability, fully characterized
LCL identifiability, fully characterized

model is identifiable from dataset \( \mathcal{D} \) if no two parameter values result in the same probability distribution
LCL identifiability, fully characterized

model is *identifiable* from dataset $\mathcal{D}$ if no two parameter values result in the same probability distribution

→ important for inference and interpretation
LCL identifiability, fully characterized

model is \textit{identifiable} from dataset $\mathcal{D}$ if no two parameter values result in the same probability distribution

$\rightarrow$ important for inference and interpretation

\textit{Theorem 1.} A $d$-feature linear context logit is identifiable from a dataset $\mathcal{D}$ if and only if

$$\text{span}\left\{\begin{bmatrix} x_C \\ 1 \end{bmatrix} \otimes (x_i - x_C) \mid C \in \mathcal{C}_\mathcal{D}, i \in C\right\} = \mathbb{R}^{d^2+d}.$$ (6)

($\mathcal{C}_\mathcal{D}$: unique choice sets in $\mathcal{D}$, $\otimes$: Kronecker product)
LCL identifiability, fully characterized

model is \textit{identifiable} from dataset $\mathcal{D}$ if no two parameter values result in the same probability distribution

$\rightarrow$ important for inference and interpretation

\textit{Theorem 1.} A $d$-feature linear context logit is identifiable from a dataset $\mathcal{D}$ if and only if

\[
\text{span} \left\{ \begin{bmatrix} x_C \\ 1 \end{bmatrix} \otimes (x_i - x_C) \mid C \in C_\mathcal{D}, i \in C \right\} = \mathbb{R}^{d^2+d}. \quad (6)
\]

($C_\mathcal{D}$: unique choice sets in $\mathcal{D}$, $\otimes$: Kronecker product)

\textit{Intuition:} need varied choice sets containing varied items
LCL extension: *Decomposed LCL (DLCL)*
LCL extension: *Decomposed LCL (DLCL)*

→ combines *mixed logit* with LCL

→ more flexible but harder to train (*expectation-maximization*)
LCL extension: *Decomposed LCL (DLCL)*

→ combines *mixed logit* with LCL

→ more flexible but harder to train (*expectation-maximization*)

\[
Pr(i \mid C) = \sum_{k=1}^{d} \pi_k \frac{\exp([B_k + A_k(x_C)_k]^T x_i)}{\sum_{j \in C} \exp([B_k + A_k(x_C)_k]^T x_j)}
\]
LCL extension: *Decomposed LCL (DLCL)*

→ combines *mixed logit* with LCL

→ more flexible but harder to train (*expectation-maximization*)

\[
\Pr(i \mid C) = \sum_{k=1}^{d} \pi_k \frac{\exp([B_k + A_k(x_C)_k]^T x_i)}{\sum_{j \in C} \exp([B_k + A_k(x_C)_k]^T x_j)}
\]

→ see paper for details
Results on choice data
Choice datasets

### Table 1: General choice datasets summary.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Choices</th>
<th>Features</th>
<th>Largest Choice Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISTRICT</td>
<td>5376</td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>DISTRICT-SMART</td>
<td>5376</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>SUSHI</td>
<td>5000</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>EXPEDIA</td>
<td>276593</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>CAR-A</td>
<td>2675</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>CAR-B</td>
<td>2206</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>CAR-ALT</td>
<td>4654</td>
<td>21</td>
<td>6</td>
</tr>
</tbody>
</table>
We consider the set of search results to be the choice set and the study a collection of temporal social network datasets, where the context effects across thirteen datasets. In addition to providing finding common patterns across datasets, we have timestamps on each edge and an edge may be observed many times (e.g., in an email network, we do not know from the network data whose out-neighbors of a node is, but not the other node's in-neighbors). Since we do not know from the network data whose will send a message to whom, we assume that such a message is sent uniformly at random through which the triangle is closed at least one time. (One could model the intermediary selection of which node sends the message to whom. However, this is outside the scope of this paper.)

Recent work and the more general social network study is also of interest for insight into sociological processes and highlights how our models can be applied to a particular domain. The set from which choices are evaluated on the held-out test set. We use...

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Choices</th>
<th>Features</th>
<th>Largest Choice Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISTRICT</td>
<td>5376</td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>DISTRICT-SMART</td>
<td>5376</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>SUSHI</td>
<td>5000</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>EXPEDIA</td>
<td>276593</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>CAR-A</td>
<td>2675</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>CAR-B</td>
<td>2206</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>CAR-ALT</td>
<td>4654</td>
<td>21</td>
<td>6</td>
</tr>
</tbody>
</table>
### Choice datasets

#### Table 1: General choice datasets summary.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Choices</th>
<th>Features</th>
<th>Largest Choice Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISTRICT</td>
<td>5376</td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>DISTRICT-SMART</td>
<td>5376</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>SUSHI</td>
<td>5000</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>EXPEDIA</td>
<td>276593</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>CAR-A</td>
<td>2675</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>CAR-B</td>
<td>2206</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>CAR-ALT</td>
<td>4654</td>
<td>21</td>
<td>6</td>
</tr>
</tbody>
</table>

- **favorite sushi types**
- **hotel bookings**
LCL improves model fit

whole-dataset negative log-likelihood (lower = better)

<table>
<thead>
<tr>
<th></th>
<th>CL</th>
<th>LCL</th>
<th>Mixed logit</th>
<th>DLCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISTRICT</td>
<td>3313</td>
<td>3130</td>
<td>3258</td>
<td>3206</td>
</tr>
<tr>
<td>DISTRICT-SMART</td>
<td>3426</td>
<td>3278*</td>
<td>3351</td>
<td>3303†</td>
</tr>
<tr>
<td>EXPEDIA</td>
<td>839505</td>
<td>837649*</td>
<td>839055</td>
<td>837569†</td>
</tr>
<tr>
<td>SUSHI</td>
<td>9821</td>
<td>9773*</td>
<td>9793</td>
<td>9764</td>
</tr>
<tr>
<td>CAR-A</td>
<td>1702</td>
<td>1694</td>
<td>1696</td>
<td>1692</td>
</tr>
<tr>
<td>CAR-B</td>
<td>1305</td>
<td>1295</td>
<td>1297</td>
<td>1284</td>
</tr>
<tr>
<td>CAR-ALT</td>
<td>7393</td>
<td>6733*</td>
<td>7301</td>
<td>7011†</td>
</tr>
</tbody>
</table>

*significant likelihood ratio test vs MNL ($p < 0.001$)
†significant likelihood ratio test vs mixed logit ($p < 0.001$)
LCL can improve out-of-sample prediction performance

Figure 2: Mean relative rank of predictions on held-out test data (lower is better). Error bars show standard error of the mean.
LCL can test individual effects for significance
LCL can test individual effects for significance

Compute std. errs. (and z-scores) for each parameter estimate using MLE asymptotic normality
LCL can test individual effects for significance

Compute std. errs. (and z-scores) for each parameter estimate using MLE asymptotic normality

<table>
<thead>
<tr>
<th>Effect (q on p)</th>
<th>$A_{pq}$ (std. err.)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>popularity on popularity</td>
<td>$-0.28$ ($0.15$)</td>
<td>0.066</td>
</tr>
<tr>
<td>availability on is maki</td>
<td>0.24 ($0.14$)</td>
<td>0.087</td>
</tr>
<tr>
<td>oiliness on oiliness</td>
<td>$-0.20$ ($0.08$)</td>
<td>0.0089</td>
</tr>
<tr>
<td>popularity on availability</td>
<td>0.19 ($0.14$)</td>
<td>0.16</td>
</tr>
<tr>
<td>availability on oiliness</td>
<td>$-0.18$ ($0.10$)</td>
<td>0.064</td>
</tr>
</tbody>
</table>
LCL can test individual effects for significance

Compute std. errs. (and z-scores) for each parameter estimate using MLE asymptotic normality

<table>
<thead>
<tr>
<th>Table 4: Five largest context effects in SUSHI.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect (q on p)</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>popularity on popularity</td>
</tr>
<tr>
<td>availability on is maki</td>
</tr>
<tr>
<td>oiliness on oiliness</td>
</tr>
<tr>
<td>popularity on availability</td>
</tr>
<tr>
<td>availability on oiliness</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5: Five largest context effects in EXPEDIA.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect (q on p)</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>location score on price</td>
</tr>
<tr>
<td>on promotion on price</td>
</tr>
<tr>
<td>review score on price</td>
</tr>
<tr>
<td>star rating on price</td>
</tr>
<tr>
<td>price on star rating</td>
</tr>
</tbody>
</table>
Social network application
What factors drive edge formation?
What factors drive edge formation?

*Preferential attachment*

(Barabási & Albert, *Science* 1999)
What factors drive edge formation?

**Preferential attachment**
(Barabási & Albert, *Science* 1999)

**Homophily**
(McPherson et al., *Annual Review of Sociology* 2001)
(Papadopoulos et al., *Nature* 2012)
What factors drive edge formation?

**Preferential attachment**
(Barabási & Albert, Science 1999)

**Homophily**
(McPherson et al., Annual Review of Sociology 2001)
(Papadopoulos et al., Nature 2012)

**Triadic closure**
(Rapoport, Bulletin of Mathematical Biophysics 1953)
(Jin et al., Physical Review E 2001)
“Choosing to grow a graph”
(Overgoor et al., SINM ’19 & WWW ’19)
(Gupta & Porter, arXiv 2020)
“Choosing to grow a graph”
(Overgoor et al., SINM ’19 & WWW ’19)
(Gupta & Porter, arXiv 2020)

so far:

chooser choice set
“Choosing to grow a graph”
(Overgoor et al., SINM ’19 & WWW ’19)
(Gupta & Porter, arXiv 2020)

so far:

chooser  
choice set

in network growth:

chooser  
choice set
“Choosing to grow a graph”

(Overgoor et al., SINM '19 & WWW '19)
(Gupta & Porter, arXiv 2020)

so far:

So far, we have chosen a set of objects: an apple, lemon, grapes, and a strawberry. These choices form a meaningful choice set.

in network growth:

In network growth, the process of choosing objects can be visualized as a graph where the chooser selects from a set of available objects. This process allows us to infer the relative importance of edge formation mechanisms from data.

Key usage

Timestamped edges → meaningful choice sets

Infer relative importance of edge formation mechanisms from data
“Choosing to grow a graph”

(Overgoor et al., *SINM* '19 & *WWW* '19)

(Gupta & Porter, arXiv 2020)

so far:

chooser choice set

in network growth:

chooser choice set

feature context effects:

Key usage

Timestamped edges → meaningful choice sets

Infer relative importance of edge formation mechanisms from data
Choosing to close triangles

*Triadic closure* offers small choice sets

→ tractable inference
→ varied choice sets
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

Our data
Timestamped edges (including repeats)
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

**Our data**
Timestamped edges (including repeats)
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

Our data
Timestamped edges (including repeats)
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

**Our data**
Timestamped edges (including repeats)
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

Our data
Timestamped edges
(including repeats)
Choosing to close triangles

**Triadic closure** offers small choice sets
→ tractable inference
→ varied choice sets

**Our data**
Timestamped edges
(including repeats)
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

**Our data**
Timestamped edges
(including repeats)
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

**Our data**
Timestamped edges (including repeats)
Choosing to close triangles

**Triadic closure** offers small choice sets
→ tractable inference
→ varied choice sets

**Our data**
Timestamped edges (including repeats)
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

**Our data**
Timestamped edges (including repeats)
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

Our data
Timestamped edges (including repeats)
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

Our data
Timestamped edges
(including repeats)
Choosing to close triangles

**Triadic closure** offers small choice sets → tractable inference → varied choice sets

**Our data**
Timestamped edges (including repeats)

![Graph diagram with nodes and arrows representing the closure of triangles.]

- **chooser**: u
- **choice set**: \{w_1, w_2, w_3\}
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

**Our data**
Timestamped edges (including repeats)

---

 chooser  
  \[ u \]  

 choice set  
  \{ w_1, w_2, w_3 \}  

 choice  
  \[ w_1 \]  

---
Choosing to close triangles

*Triadic closure* offers small choice sets
→ tractable inference
→ varied choice sets

Node features
1. in-degree of $w$
2. # shared neighbors of $u, w$
3. weight of edge $w \rightarrow u$
4. time since last edge into $w$
5. time since last edge out of $w$
6. time since last $w \rightarrow u$ edge

Our data
Timestamped edges (including repeats)

** chooser**
$u$

** choice set**
$\{w_1, w_2, w_3\}$

** choice**
$w_1$
Context matters in triadic closure

Datasets
- email-enron
- email-eu
- email-w3c
- wiki-talk
- reddit-hyperlink
- bitcoin-alpha
- bitcoin-otc
- mathoverflow
- college-msg
- facebook-wall
- sms-a
- sms-b
- sms-c
Context matters in triadic closure

Datasets
- email-enron
- email-eu
- email-w3c
- wiki-talk
- reddit-hyperlink
- bitcoin-alpha
- bitcoin-otc
- mathoverflow
- college-msg
- facebook-wall
- sms-a
- sms-b
- sms-c
Context matters in triadic closure

Datasets:
- email-enron
- email-eu
- email-w3c
- wiki-talk
- reddit-hyperlink
- bitcoin-alpha
- bitcoin-otc
- mathoverflow
- college-msg
- facebook-wall
- sms-a
- sms-b
- sms-c

Synthetic data, no context effects
Context matters in triadic closure

Datasets
- email-enron
- email-eu
- email-w3c
- wiki-talk
- reddit-hyperlink
- bitcoin-alpha
- bitcoin-otc
- mathoverflow
- college-msg
- facebook-wall
- sms-a
- sms-b
- sms-c

Synthetic data, no context effects

Commenting network, linear context effects
Context matters in triadic closure

Datasets
- email-enron
- email-eu
- email-w3c
- wiki-talk
- reddit-hyperlink
- bitcoin-alpha
- bitcoin-otc
- mathoverflow
- college-msg
- facebook-wall
- sms-a
- sms-b
- sms-c

Synthetic data, no context effects

Commenting network, linear context effects

Email network, nonlinear context effects?
LCL reveals interpretable feature context effects
LCL reveals interpretable feature context effects
LCL reveals interpretable feature context effects

Node features
(left-right, top-bottom)

1. in-degree
2. shared neighbors
3. reciprocal weight
4. send recency
5. receive recency
6. reciprocal recency

context effect matrix $A$
red: +, blue: -, white: 0
(column acts on row)
LCL reveals interpretable feature context effects

Node features
(left-right, top-bottom)
1. in-degree
2. shared neighbors
3. reciprocal weight
4. send recency
5. receive recency
6. reciprocal recency

context effect matrix $A$
red: +, blue: -, white: 0
(column acts on row)

increasing $L_1$ regularization on $A$

synthetic-mnl

$\lambda = 0 \quad 0.005 \quad 0.01 \quad 0.05 \quad 0.1$
LCL reveals interpretable feature context effects

context effect matrix $A$
red: +, blue: -, white: 0
(column acts on row)

increasing $L_1$ regularization on $A$

Node features (left-right, top-bottom)
1. in-degree
2. shared neighbors
3. reciprocal weight
4. send recency
5. receive recency
6. reciprocal recency

NLL (lower = better fit)
LCL reveals interpretable feature context effects

Node features (left-right, top-bottom)
1. in-degree
2. shared neighbors
3. reciprocal weight
4. send recency
5. receive recency
6. reciprocal recency

context effect matrix $A$
red: +, blue: -, white: 0 (column acts on row)

increasing $L_1$ regularization on $A$

NLL (lower = better fit)

$p = 0.001$ LRT threshold
LCL reveals interpretable feature context effects

Node features (left-right, top-bottom)
1. in-degree
2. shared neighbors
3. reciprocal weight
4. send recency
5. receive recency
6. reciprocal recency

context effect matrix $A$
red: +, blue: -, white: 0
(column acts on row)

increasing $L_1$ regularization on $A$

NLL (lower = better fit)

$p = 0.001$ LRT threshold
LCL reveals interpretable feature context effects

Node features (left-right, top-bottom)
1. in-degree
2. shared neighbors
3. reciprocal weight
4. send recency
5. receive recency
6. reciprocal recency

context effect matrix $A$
red: +, blue: -, white: 0 (column acts on row)

Increasing $L_1$ regularization on $A$

NLL (lower = better fit)

$p = 0.001$ LRT threshold
LCL reveals interpretable feature context effects

context effect matrix $A$
red: +, blue: -, white: 0
(column acts on row)

Node features
(left-right, top-bottom)
1. in-degree
2. shared neighbors
3. reciprocal weight
4. send recency
5. receive recency
6. reciprocal recency

increasing $L_1$ regularization on $A$

NLL (lower = better fit)

$p = 0.001$ LRT threshold

“cluttered inbox”
high choice set reciprocal recency → in-degree less important
LCL reveals interpretable feature context effects

context effect matrix $A$
red: +, blue: -, white: 0
(column acts on row)

increasing $L_1$ regularization on $A$

Node features
(left-right, top-bottom)
1. in-degree
2. shared neighbors
3. reciprocal weight
4. send recency
5. receive recency
6. reciprocal recency

$\lambda = 0$ 0.005 0.01 0.05 0.1

$\text{NLL (lower = better fit)}$
$p = 0.001$ LRT threshold

“cluttered inbox”
high choice set reciprocal recency
$\rightarrow$ in-degree less important

red: “cocktail party introduction”
high choice set in-degree
$\rightarrow$ shared neighbors more important

blue: “familiarity saturation”
high choice set shared neighbors
$\rightarrow$ shared neighbors less important
Concluding thoughts

Key takeaways
*Feature context effects* extend item-level effects
LCL offers an interpretable and tractable way to reveal them

Future work
- Non-linear context effects
- Negative sampling
- Discovering relational effects

Causal context effects?
See our other KDD ’21 paper:
“Choice Set Confounding in Discrete Choice”

Thank you!
More questions or ideas?
Email me: kt@cs.cornell.edu
@kiran_tomlinson

Acknowledgments
Funding from NSF, ARO
Thanks to Johan Ugander, Jan Overgoor, and Sophia Franco