Learning Interpretable Feature Context Effects in Discrete Choice

Kiran Tomlinson

@ Stanford SOAL

research with Austin R. Benson

Code: bit.ly/lcl-code
Data: bit.ly/lcl-data
1. 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
“The fundamental problem of discrete choice”
“The fundamental problem of discrete choice”

choice set

choice

...
The classic model: \textit{multinomial logit (MNL)}

(McFadden, \textit{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: *multinomial logit (MNL)*

(McFadden, *Frontiers in Econometrics* 1973)

Assume item \( i \) has utility \( u_i \)

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

\[\begin{array}{cccc}
C & \text{Apple} & \text{Lemon} & \text{Grapes} & \text{Strawberry} \\
\hline
u_i & 1 & -1 & 0 & 2 \\
Pr(i \mid C) & .24 & .03 & .09 & .64
\end{array}\]
The classic model: *multinomial logit (MNL)*

(McFadden, *Frontiers in Econometrics* 1973)

Assume item \( i \) has utility \( u_i \)

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

\( C \)

\[
\begin{array}{cccc}
1 & -1 & 0 & 2 \\
\Pr(i \mid C) & .24 & .03 & .09 & .64
\end{array}
\]

Econometric derivation:
- Draw random utilities \( u_i + \epsilon_i \) for each item (\( \epsilon_i \) i.i.d. Gumbel)
- Rational agent picks \( \arg\max_{i \in C} u_i + \epsilon_i \)

\( \rightarrow \) MNL
The classic model: **multinomial logit (MNL)**

(McFadden, *Frontiers in Econometrics* 1973)

Assume *item* \( i \) has *utility* \( u_i \)

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

Econometric derivation:
- Draw random utilities \( u_i + \epsilon_i \) for each item (\( \epsilon_i \) i.i.d. Gumbel)
- Rational agent picks \( \arg\max_{i \in C} u_i + \epsilon_i \)

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

\[
\frac{\Pr(i \mid C)}{\Pr(j \mid C)} = \frac{\Pr(i \mid C')}{\Pr(j \mid C')}
\]

(Luce, *Individual Choice Behavior* 1959)
Learning an MNL from choice data
Learning an MNL from choice data

**Likelihood** of utilities $u$ given dataset $\mathcal{D}$:

$$L(u; \mathcal{D}) = \prod_{(i,C) \in \mathcal{D}} \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$
Learning an MNL from choice data

Likelihood of utilities $u$ given dataset $\mathcal{D}$:

$$L(u; \mathcal{D}) = \prod_{(i, C) \in \mathcal{D}} \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}$$

Log-likelihood is concave:

$$\ell(u; \mathcal{D}) = \sum_{(i, C) \in \mathcal{D}} u_i - \log \sum_{j \in C} \exp(u_j)$$
Learning an MNL from choice data

Likelihood of utilities $u$ given dataset $\mathcal{D}$: 

$$
\mathcal{L}(u; \mathcal{D}) = \prod_{(i,C) \in \mathcal{D}} \frac{\exp(u_i)}{\sum_{j \in C} \exp(u_j)}
$$

Log-likelihood is concave: 

$$
\ell(u; \mathcal{D}) = \sum_{(i,C) \in \mathcal{D}} u_i - \log \sum_{j \in C} \exp(u_j)
$$

$\rightarrow$ gradient descent to learn $u$

(SGD, Adam, …)
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)

**Asymmetric dominance**  
(Huber et al., 1982)

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

- $10
- $15
- $20
- $25
Problem for MNL: context effects

The choice set influences preferences.

**Compromise**
(Simonson, 1989)

$10  
$15  
$20  
$25

**Asymmetric dominance**
(Huber et al., 1982)

![Economist.com Subscriptions](image)

Welcome to The Economist Subscription Centre

Pick the type of subscription you want to buy or renew.

- Economist.com subscription - US $59.00
- Print subscription - US $125.00
- Print & web subscription - US $125.00

One-year subscription to Economist.com.
Includes online access to all articles from The Economist since 1997.

One-year subscription to the print edition of The Economist.

One-year subscription to the print edition of The Economist and online access to all articles from The Economist since 1997.
Problem for MNL: context effects

The choice set influences preferences.

**Compromise**
(Simonson, 1989)

**Asymmetric dominance**
(Huber et al., 1982)
Problem for MNL: *context effects*

The choice set influences preferences.

**Compromise**  
(Simonson, 1989)

**Asymmetric dominance**  
(Huber et al., 1982)

**Similarity**  
(Tversky, 1972)

(Ariely, 2008)
Problem for MNL: context effects

The choice set influences preferences.

**Compromise**  
(Simonson, 1989)

**Asymmetric dominance**  
(Huber et al., 1982)

**Similarity**  
(Tversky, 1972)

---

- **David Perdue**  
  Rep.  
  2,462,617  
  49.7%

- **Jon Ossoff**  
  Dem.  
  2,374,519  
  47.9%
Problem for MNL: **context effects**

The choice set influences preferences.

**Compromise**  
(Simonson, 1989)

**Asymmetric dominance**  
(Huber et al., 1982)

**Similarity**  
(Tversky, 1972)

(Ariely, 2008)

(Ariely, 2008)
**Problem for MNL: context effects**

The choice set influences preferences.

**Compromise**  
(Simonson, 1989)

**Asymmetric dominance**  
(Huber et al., 1982)

**Similarity**  
(Tversky, 1972)

IIA violations:

\[
\frac{\Pr(i \mid C)}{\Pr(j \mid C)} \neq \frac{\Pr(i \mid C')}{\Pr(j \mid C')}
\]

(Ariely, 2008)
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:

$$
\text{Pr}(i \mid C) = \frac{\exp \left( \sum_{k \in C \setminus i} u_{ik} \right)}{\sum_{j \in C} \exp \left( \sum_{k \in C \setminus i} u_{jk} \right)}
$$
Natural context effect model: \textbf{CDM}  
(Seshadri, Peysakhovich, & Ugander, ICML 2019)

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

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

\begin{itemize}
\item $u_{\text{Loeffler, Collins}} < 0$
\item $u_{\text{Collins, Loeffler}} < 0$
\end{itemize}
Natural context effect model: CDM

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

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

$u_{\text{Loeffler, Collins}} < 0$

$u_{\text{Collins, Loeffler}} < 0$

$u_{\text{Print&Web, Print}} > 0$)
Natural context effect model: \textbf{CDM}  
(Seshadri, Peysakhovich, & Ugander, ICML 2019)

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

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

Assumes no higher-order effects (i.e., context effects decompose additively into effects of items)
2. 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:** comedy,  
  *in_top_10:* False,  
  *has_new_episodes:* False,  
  *producer:* NBC

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

- **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}

Feature vector $x_i \in \mathbb{R}^d$ for each item $i$
Preference vector $\theta \in \mathbb{R}^d$
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: \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:
$$\Pr(i \mid C) = \frac{\exp(\theta^T x_i)}{\sum_{j \in C} \exp(\theta^T x_j)}$$

Preference coefficient $\theta_k$ is easy to interpret: importance of the $k^{\text{th}}$ 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: \( u_i = \theta^T x_i \)  
Contextual utility: \( u_{i,C} = [\theta + F(C)]^T x_i \)
Incorporating *feature context effects* into conditional logit

Conditional logit utility: $u_i = \theta^T x_i$  \quad \rightarrow \quad \text{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 \text{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 \( A \in \mathbb{R}^{d \times d} \)

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

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

\[\rightarrow \text{convex negative log-likelihood}\]

\[\rightarrow \theta: \text{base preference coefficients}\]

\[\rightarrow A_{pq} > 0: \text{when } q \text{ is high in the choice set, } p \text{ is more preferred}\]

\[\rightarrow A_{pq} < 0: \text{when } q \text{ is high in the choice set, } p \text{ is less preferred}\]
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
LCL example: restaurant selection

item features:
- price
- service speed
- wine selection

mean choice set price

Icons by icons8.com
LCL example: restaurant selection

item features:
- price
- service speed
- wine selection

importance of:
- speed
- wine selection

mean choice set price

A

\[
\begin{bmatrix}
0 & 0 & 0 \\
-1 & 0 & 0 \\
1 & 0 & 0
\end{bmatrix}
\]
Linear feature context effects appear in real data
3. Decomposed LCL
Motivation: the problem of intercepts
Motivation: the problem of intercepts
Motivation: the problem of intercepts

what happens with mean vector zero?
Motivation: the problem of intercepts

what happens with mean vector zero?

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

sub-model for the context effect exerted by each feature
One option: \textit{Decomposed Linear Context Logit (DLCL)}

\textit{sub-model} for the context effect exerted by each feature

\[ u_{i,C,k} : \text{utility of } i \text{ in } C \text{ in the context of feature } k \]

\[ u_{i,C,k} = [B_k + F_k(C)]^T x_i \]

(matrix subscript = column)
One option: *Decomposed Linear Context Logit (DLCL)*

*sub-model* for the context effect exerted by each feature

→ $u_{i,C,k}$: utility of $i$ in $C$ in the context of feature $k$

$$u_{i,C,k} = [B_k + F_k(C)]^T x_i$$

→ $B$: matrix of intercepts

(matrix subscript = column)
One option: *Decomposed Linear Context Logit (DLCL)*

**sub-model** for the context effect exerted by each feature

→ $u_{i,C,k}$: utility of $i$ in $C$ in the context of feature $k$

$$u_{i,C,k} = [B_k + F_k(C)]^T x_i$$

→ $B$: matrix of intercepts

→ LCL assumptions yields $F_k(C) = A_k(x_C)_k$
One option: *Decomposed Linear Context Logit (DLCL)*

**Sub-model** for the context effect exerted by each feature

\[ u_{i,C,k} = [B_k + F_k(C)]^T x_i \]

- \( u_{i,C,k} \): utility of \( i \) in \( C \) in the context of feature \( k \)
- \( B \): matrix of intercepts
- LCL assumptions yield \( F_k(C) = A_k(x_C)_k \)
- Combine sub-models in mixture (w/ proportions \( \pi \))
One option: **Decomposed Linear Context Logit (DLCL)**

sub-model for the context effect exerted by each feature

→ $u_{i,C,k}$: utility of $i$ in $C$ in the context of feature $k$

$$u_{i,C,k} = [B_k + F_k(C)]^T x_i$$

→ $B$: matrix of intercepts

→ LCL assumptions yields $F_k(C) = A_k(x_{C})_k$

→ combine sub-models in mixture (w/ proportions $\pi$)

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

(matrix subscript = column)
4. Identifiability and estimation
LCL identifiability, fully characterized
LCL identifiability, fully characterized

model is \textit{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

→ important for inference and interpretation

\begin{align}
\text{Theorem 1. A } d\text{-feature linear context logit is identifiable from a dataset } \mathcal{D} \text{ 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}. \quad (6)
\end{align}

($\mathcal{C}_\mathcal{D}$: 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

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

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

\textit{intuition}: need varied choice sets containing varied items
LCL identifiability, more intuition
LCL identifiability, more intuition

Proposition 2. No $d$-feature linear context logit is identifiable from a dataset $\mathcal{D}$ if it does not include a set of $d + 1$ choice sets with affinely independent mean feature vectors.
LCL identifiability, more intuition

**Proposition 2.** No $d$-feature linear context logit is identifiable from a dataset $\mathcal{D}$ if it does not include a set of $d + 1$ choice sets with affinely independent mean feature vectors.

**Proposition 3.** If a dataset contains $d + 1$ distinct choice sets $C_0, \ldots, C_d$ such that

i. the set of mean feature vectors $\{x_{C_0}, \ldots, x_{C_d}\}$ is affinely independent (the necessary condition from Proposition 2) and

ii. in each choice set $C_i$, there is some set of $d + 1$ items with affinely independent features,

then we can uniquely identify a $d$-feature LCL.
Mixture models are rough

see:


LCL and DLCL Estimation
LCL and DLCL Estimation

LCL

negative log-likelihood convex, so use gradient descent
LCL and DLCL Estimation

LCL

negative log-likelihood convex, so use gradient descent

$$-\ell(\theta, A; \mathcal{D}) = \sum_{(i,C) \in \mathcal{D}} -\left(\theta + Ax_C\right)^T x_i + \log \sum_{j \in C} \exp \left(\left[\theta + Ax_C\right]^T x_j\right)$$
LCL and DLCL Estimation

**LCL**

negative log-likelihood convex, so use gradient descent

$$-\ell(\theta, A; D) = \sum_{(i, C) \in D} - (\theta + Ax_C)^T x_i + \log \sum_{j \in C} \exp \left( [\theta + Ax_C]^T x_j \right)$$

**DLCL**

log-likelihood not convex…

but, *expectation-maximization (EM)* algorithm performs well

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Adam NLL</th>
<th>EM NLL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISTRICT</td>
<td>3206</td>
<td>3041</td>
</tr>
<tr>
<td>DISTRICT-SMART</td>
<td>3303</td>
<td>3144</td>
</tr>
<tr>
<td>EXPEDIA</td>
<td>837569</td>
<td>805055</td>
</tr>
<tr>
<td>SUSHI</td>
<td>9764</td>
<td>9709</td>
</tr>
<tr>
<td>CAR-A</td>
<td>1692</td>
<td>1684</td>
</tr>
<tr>
<td>CAR-B</td>
<td>1284</td>
<td>1246</td>
</tr>
<tr>
<td>CAR-ALT</td>
<td>7011</td>
<td>6369</td>
</tr>
</tbody>
</table>
5. Results on choice data
We consider the set of search results to be the choice set and the
analysis of six choice datasets coming from online and survey
with geometric properties as features. Survey respondents were
studied on a collection of temporal social network datasets, where the
general choice datasets above come with their own specialized
effects across thirteen datasets. In addition to providing
model in all cases. However, the
ti
ability in
We do not do this here.
features) components for mixed logit to provide a fair comparison
features, this social network study is also of
We focus speci-
choice by the node

<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 worth studying for their insight into human behavior as well as for with geometric properties as features. Survey respondents were which each respondent ranked 10 sushi (randomly selected from favorite sushi types).

The general choice datasets above come with their own specialized properties of the nodes. This setting allows us to examine comparable study a collection of temporal social network datasets, where the effects across datasets is one key step in showing that these e

ti

ters across them. However,

In all datasets, we standardize the features to have zero mean and unit variance, which allows us to more meaningfully compare alternative-fuel vehicles (e.g., electric, compressed natural gas).

The authors of [dataset as “good predictors of compactness. “ The

The dataset is identical, but contains the subset of features identi

understanding of anti-gerrymandering laws). The

pairwise comparisons between US congressional district shapes,

contains user searches, displayed results, and which hotel was booked. We consider a set of 100 options) from favorite to least favorite. We consider data (Table 1). The classic

6.1 General Choice Datasets

against DLCL (which always uses features) components for mixed logit to provide a fair comparison 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>

Our net-

...
Choice datasets

- favorite sushi types (random choice sets in survey)
- hotel bookings (choice sets = search results)

<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 context effects we apply (in the form of Adam weight decay) identify learned parameters across datasets. However, the authors of [ef{27}] comes from online hotel booking. It contains 276593 options and was closed (note that with probability 0.01, we do not do this here.). With this setup, we can reconstruct choice sets for each triangle closure with probability 0.59, uniformly at random through which the triangle closure coexists, or decides to close a triangle with probability 0.90. This is the same setup used by the Jackson–Rogers model and real-world networks show evidence of this phenomenon.

<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>
**LCL improves model fit**

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

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

Mean (std. dev) relative rank (0 = perfect prediction, 1 = always wrong)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MNL</th>
<th>LCL</th>
<th>Mixed logit</th>
<th>DLCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>District</td>
<td>.3680 (.4823)</td>
<td>.3253 (.4685)</td>
<td>.3188 (.4660)</td>
<td>.3225 (.4674)</td>
</tr>
<tr>
<td>District-Smart</td>
<td>.4006 (.4900)</td>
<td>.3764 (.4845)</td>
<td>.3271 (.4692)</td>
<td>.3448 (.4753)</td>
</tr>
<tr>
<td>Expedia</td>
<td>.3859 (.2954)</td>
<td>.3800* (.2945)</td>
<td>.3201 (.2825)</td>
<td>.3195† (.2823)</td>
</tr>
<tr>
<td>Sushi</td>
<td>.2727 (.2751)</td>
<td>.2737 (.2781)</td>
<td>.2724 (.2765)</td>
<td>.2732 (.2765)</td>
</tr>
<tr>
<td>Car-A</td>
<td>.3570 (.4791)</td>
<td>.3570 (.4791)</td>
<td>.3570 (.4791)</td>
<td>.3570 (.4791)</td>
</tr>
<tr>
<td>Car-B</td>
<td>.3326 (.4711)</td>
<td>.3213 (.4670)</td>
<td>.3303 (.4703)</td>
<td>.3235 (.4678)</td>
</tr>
<tr>
<td>Car-Alt</td>
<td>.2944 (.2875)</td>
<td>.2661* (.2819)</td>
<td>.2931 (.2966)</td>
<td>.2798 (.2837)</td>
</tr>
</tbody>
</table>

*significant Wilcoxon test vs MNL (p < 0.001)
LCL can test individual effects for significance

We consider the set of search results to be the choice set and the worth studying for their insight into human behavior as well as for with geometric properties as features. Survey respondents were the general choice datasets above come with their own specialized properties of the nodes. This setting allows us to examine comparable their theoretical interest or use in prediction. To this end, we also context e set of features. For this reason, it is not possible to compare feature model in all cases.

The authors of [35] as “good predictors of compactness.” The dataset is identical, but contains the subset of features identi about the dataset features and preprocessing steps. The LCL is iden- and unit variance, which allows us to more meaningfully compare similar, but has choice sets of six hypothetical cars and focuses on sedan) and transmission (manual, automatic).

alternative-fuel vehicles (e.g., electric, compressed natural gas).

regular-
LCL can test individual effects for significance

constrain all but one entry of $A$ to 0
LCL can test individual effects for significance

constrain all but one entry of $A$ to 0

$\rightarrow$ loss still convex
LCL can test individual effects for significance

constrain all but one entry of $A$ to 0

$\rightarrow$ loss still convex

$\rightarrow$ test \textit{strength} of single effect
LCL can test individual effects for significance

constrain all but one entry of $A$ to 0
$\rightarrow$ loss still convex
$\rightarrow$ test *strength* of single effect
$\rightarrow$ constrained model nests cond. logit
LCL can test individual effects for significance

constrain all but one entry of $A$ to 0
  → loss still convex
  → test *strength* of single effect
  → constrained model nests cond. logit

$\bar{A}_{pq}$: learned param in constrained model
LCL can test individual effects for significance

\[ \text{constraint all but one entry of } A \text{ to 0} \]
\[ \rightarrow \text{loss still convex} \]
\[ \rightarrow \text{test strength of single effect} \]
\[ \rightarrow \text{constrained model nests cond. logit} \]

\[ \overline{A}_{pq} \]: learned param in constrained model

### Table 5: Five largest context effects in sushi.

<table>
<thead>
<tr>
<th>Effect (q on p)</th>
<th>( A_{pq} )</th>
<th>( \overline{A}_{pq} )</th>
<th>LRT</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>popularity on popularity</td>
<td>-0.28</td>
<td>-0.11</td>
<td>3.0</td>
<td>0.081</td>
</tr>
<tr>
<td>availability on is maki</td>
<td>0.24</td>
<td>0.04</td>
<td>0.39</td>
<td>0.53</td>
</tr>
<tr>
<td>oiliness on oiliness</td>
<td>-0.20</td>
<td>-0.26</td>
<td>23</td>
<td>( 1.5 \times 10^{-6} )</td>
</tr>
<tr>
<td>popularity on availability</td>
<td>0.19</td>
<td>0.09</td>
<td>2.3</td>
<td>0.13</td>
</tr>
<tr>
<td>availability on oiliness</td>
<td>-0.18</td>
<td>-0.04</td>
<td>0.74</td>
<td>0.39</td>
</tr>
</tbody>
</table>
LCL can test individual effects for significance

constrain all but one entry of $A$ to 0
→ loss still convex
→ test strength of single effect
→ constrained model nests cond. logit

$\bar{A}_{pq}$: learned param in constrained model
LCL can test individual effects for significance

\[ A \rightarrow \text{constrain all but one entry of } A \text{ to 0} \]
\[ \rightarrow \text{loss still convex} \]
\[ \rightarrow \text{test strength of single effect} \]
\[ \rightarrow \text{constrained model nests cond. logit} \]

\[ \overline{A}_{pq} : \text{learned param in constrained model} \]

<table>
<thead>
<tr>
<th>Table 5: Five largest context effects in SUSHI.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect (q on p)</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 6: Five largest context effects in EXPEDIA.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect (q on p)</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>

<table>
<thead>
<tr>
<th>Table 7: Five largest context effects in CAR-ALT.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect (q on p)</td>
</tr>
<tr>
<td>truck on truck</td>
</tr>
<tr>
<td>van on van</td>
</tr>
<tr>
<td>suv on station wagon</td>
</tr>
<tr>
<td>station wagon on station wagon</td>
</tr>
<tr>
<td>sports car on station wagon</td>
</tr>
</tbody>
</table>
LCL can test individual effects for significance

\[ \tilde{A}_{pq} \]: learned param in constrained model

\[ \tilde{A}_{pq} \]: learned param in constrained model

\[ A_{pq} \]: constrains all but one entry of \( A \) to 0

\[ \rightarrow \] loss still convex

\[ \rightarrow \] test strength of single effect

\[ \rightarrow \] constrained model nests cond. logit

\[ \rightarrow \] test **causal**?

<table>
<thead>
<tr>
<th>Effect (q on p)</th>
<th>( A_{pq} )</th>
<th>( \tilde{A}_{pq} )</th>
<th>LRT</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>popularity on popularity</td>
<td>-0.28</td>
<td>-0.11</td>
<td>3.0</td>
<td>0.081</td>
</tr>
<tr>
<td>availability on is maki</td>
<td>0.24</td>
<td>0.04</td>
<td>0.39</td>
<td>0.53</td>
</tr>
<tr>
<td>oiliness on oiliness</td>
<td>-0.20</td>
<td>-0.26</td>
<td>23.0</td>
<td>1.5 \times 10^{-6}</td>
</tr>
<tr>
<td>popularity on availability</td>
<td>0.19</td>
<td>0.09</td>
<td>2.3</td>
<td>0.13</td>
</tr>
<tr>
<td>availability on oiliness</td>
<td>-0.18</td>
<td>-0.04</td>
<td>0.74</td>
<td>0.39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect (q on p)</th>
<th>( A_{pq} )</th>
<th>( \tilde{A}_{pq} )</th>
<th>LRT</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>location score on price</td>
<td>-0.47</td>
<td>-0.13</td>
<td>10.0</td>
<td>0.002</td>
</tr>
<tr>
<td>on promotion on price</td>
<td>0.27</td>
<td>0.13</td>
<td>17.0</td>
<td>3.5 \times 10^{-5}</td>
</tr>
<tr>
<td>review score on price</td>
<td>-0.19</td>
<td>-0.13</td>
<td>29.0</td>
<td>8.6 \times 10^{-8}</td>
</tr>
<tr>
<td>star rating on price</td>
<td>0.15</td>
<td>0.20</td>
<td>65.0</td>
<td>6.0 \times 10^{-16}</td>
</tr>
<tr>
<td>price on star rating</td>
<td>0.10</td>
<td>0.03</td>
<td>4.0</td>
<td>0.046</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect (q on p)</th>
<th>( A_{pq} )</th>
<th>( \tilde{A}_{pq} )</th>
<th>LRT</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>truck on truck</td>
<td>1.06</td>
<td>0.83</td>
<td>239</td>
<td>&lt; 10^{-16}</td>
</tr>
<tr>
<td>van on van</td>
<td>0.94</td>
<td>0.97</td>
<td>309</td>
<td>&lt; 10^{-16}</td>
</tr>
<tr>
<td>suv on station wagon</td>
<td>0.89</td>
<td>0.98</td>
<td>-0.21</td>
<td>1.0</td>
</tr>
<tr>
<td>station wagon on station wagon</td>
<td>0.88</td>
<td>0.93</td>
<td>153</td>
<td>&lt; 10^{-16}</td>
</tr>
<tr>
<td>sports car on station wagon</td>
<td>0.86</td>
<td>0.96</td>
<td>-0.21</td>
<td>1.0</td>
</tr>
</tbody>
</table>
LCL can test individual effects for significance

constrain all but one entry of $A$ to 0

$\rightarrow$ loss still convex

$\rightarrow$ test strength of single effect

$\rightarrow$ constrained model nests cond. logit

$\overline{A}_{pq}$: learned param in constrained model

are these effects causal?

$\rightarrow$ choice set assignment (stay tuned)
6. 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)

**Fitness**
(Bianconi & Barabási, *Europhysics Letters* 2001)
(Caldarelli et al., *Physical Review Letters* 2002)

**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)

**Fitness**  
(Bianconi & Barabási, Europhysics Letters 2001)  
(Caldarelli et al., Physical Review Letters 2002)

**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., S/NM '19 & WWW '19)
(Gupta & Porter, arXiv 2020)

so far:

chooser choice set

in network growth:

chooser

choice set

**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

 chooses from:

 choice set

Key usage
Timestamped edges → meaningful choice sets
Infer relative importance of edge formation mechanisms from data

in network growth:

 chooser

 choice set

feature context effects:

 vs.

 chooser

 choice set
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
   $\rightarrow$ tractable inference
   $\rightarrow$ 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)

```
\begin{align*}
    \text{chooser} & : u \\
    \text{choice set} & : \{w_1, w_2, w_3\}
\end{align*}
```
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)

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

chooser
$u$

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

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

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 $A$

NLL (lower = better fit)

synthetic-mnl
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 $A$

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

A34 increasing 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 $A$

$\lambda = 0$ 0.005 0.01 0.05 0.1

NLL (lower = better fit)

$p = 0.001$ LRT threshold

synthetic-mnl
synthetic-lcl
email-enron
email-eu
email-w3c
facebook-wall
reddit-hyperlink

$\lambda = 0$ 0.005 0.01 0.05 0.1

+0.1% +0.05%

+4% +4%

+2% +4%

+2% +2%

+2% +2%

+2% +1%
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 $\rightarrow$ 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 $A$

NLL (lower = better fit)
$p = 0.001$ LRT threshold

"cluttered inbox"
high choice set reciprocal recency → in-degree less important

red: "cocktail party introduction"
high choice set in-degree → shared neighbors more important

blue: "familiarity saturation"
high choice set shared neighbors → shared neighbors less important

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

"synthetic-mnl" synthetic-lcl email-enron email-eu email-w3c facebook-wall reddit-hyperlink
Similar datasets have similar feature context effects

T-SNE embedding of learned $A$ matrices
Concluding thoughts
Concluding thoughts

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

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

Challenges
Features correlate
Causal context effects?
Handling nonlinearity?
Concluding thoughts

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

Challenges
Features correlate
Causal context effects?
Handling nonlinearity?

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

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