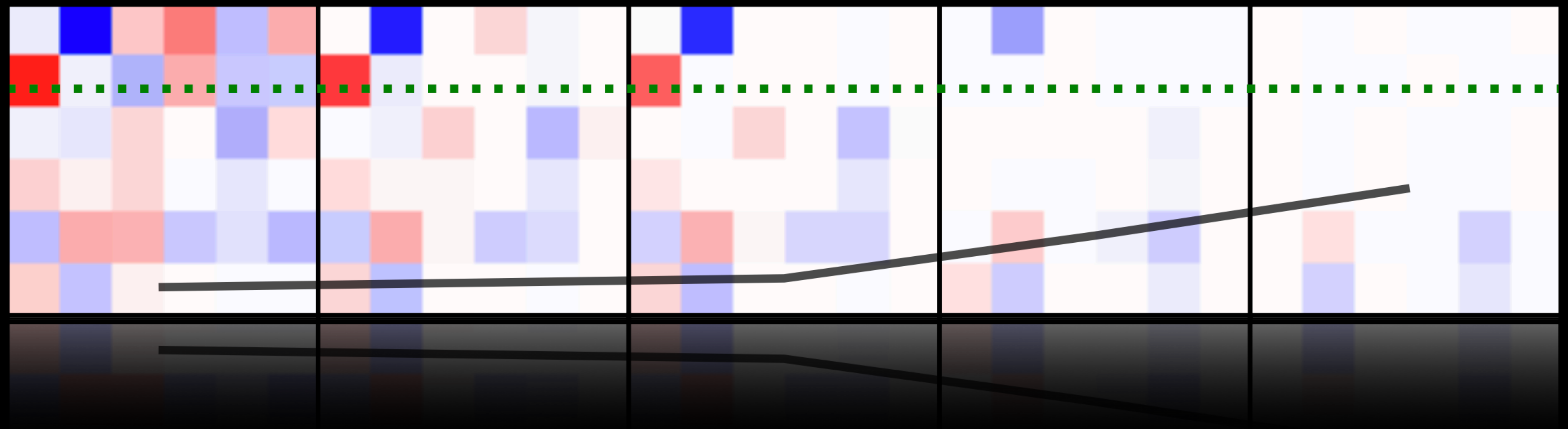
Code: bit.ly/lcl-code
Data: bit.ly/lcl-data
Slides: bit.ly/lcl-kdd-slides



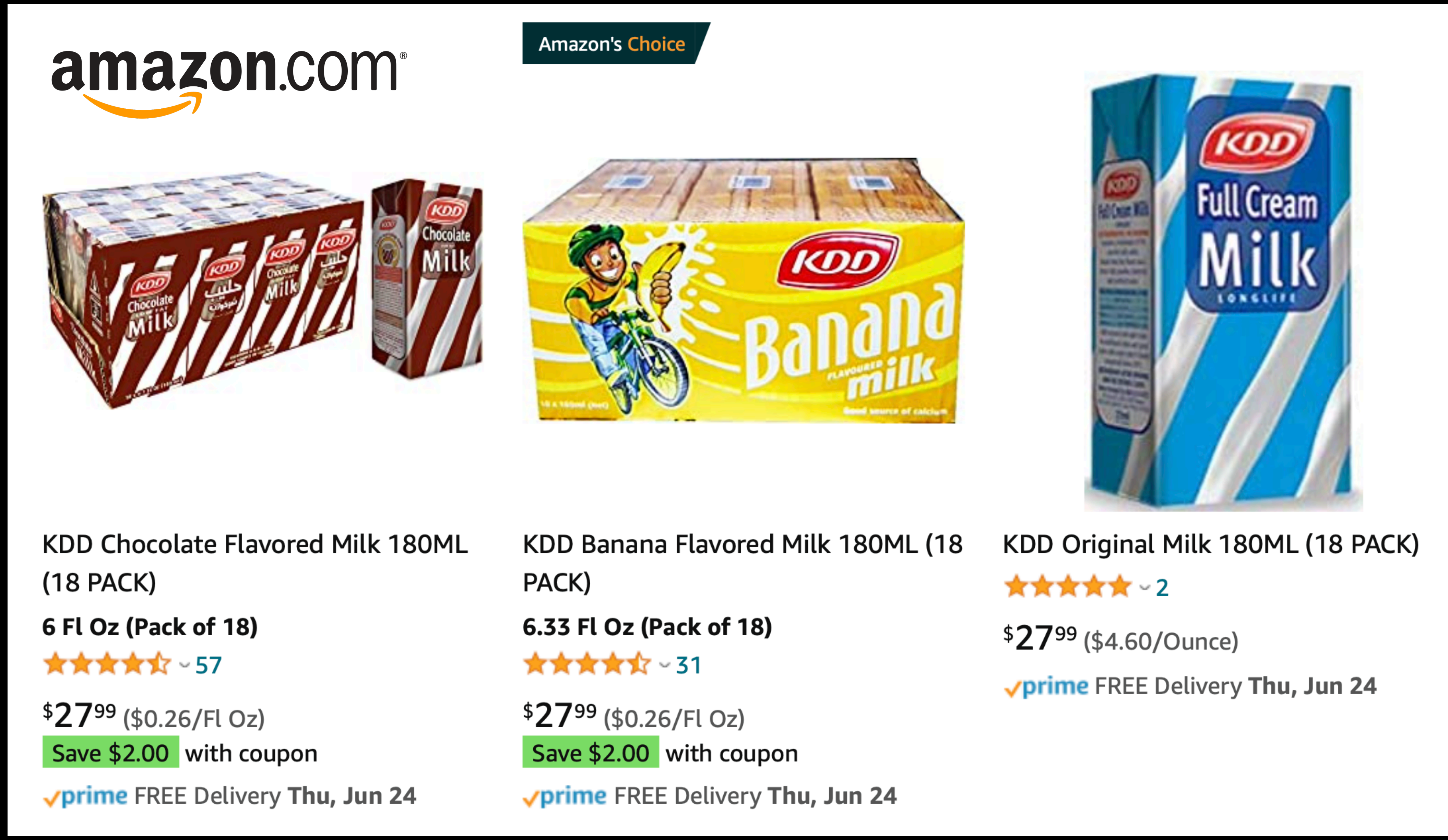
Learning Interpretable Feature Context Effects in Discrete Choice

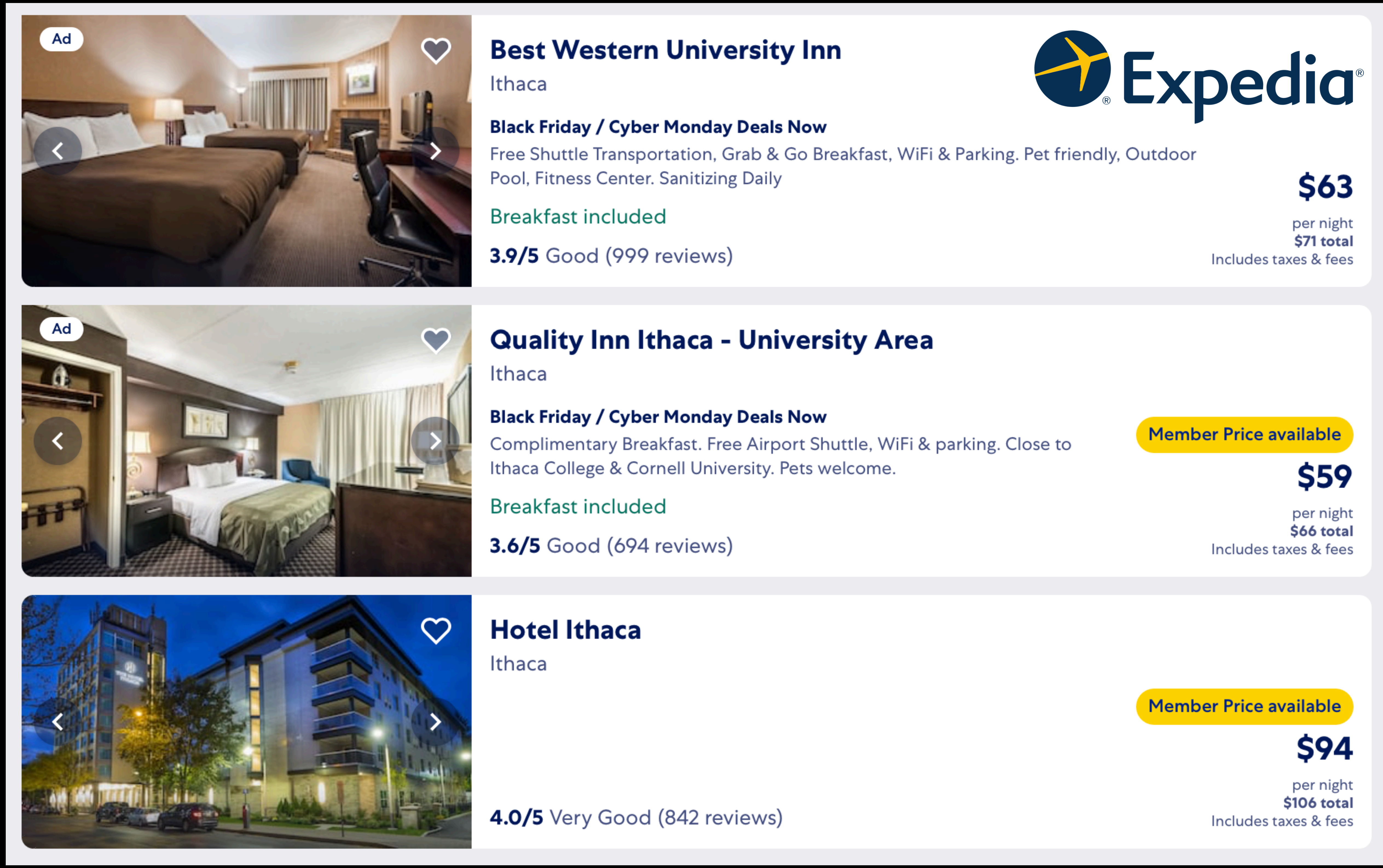
Kiran Tomlinson
PhD Student, Cornell CS

research with Austin R. Benson

Choices and context effects

Discrete choices are everywhere

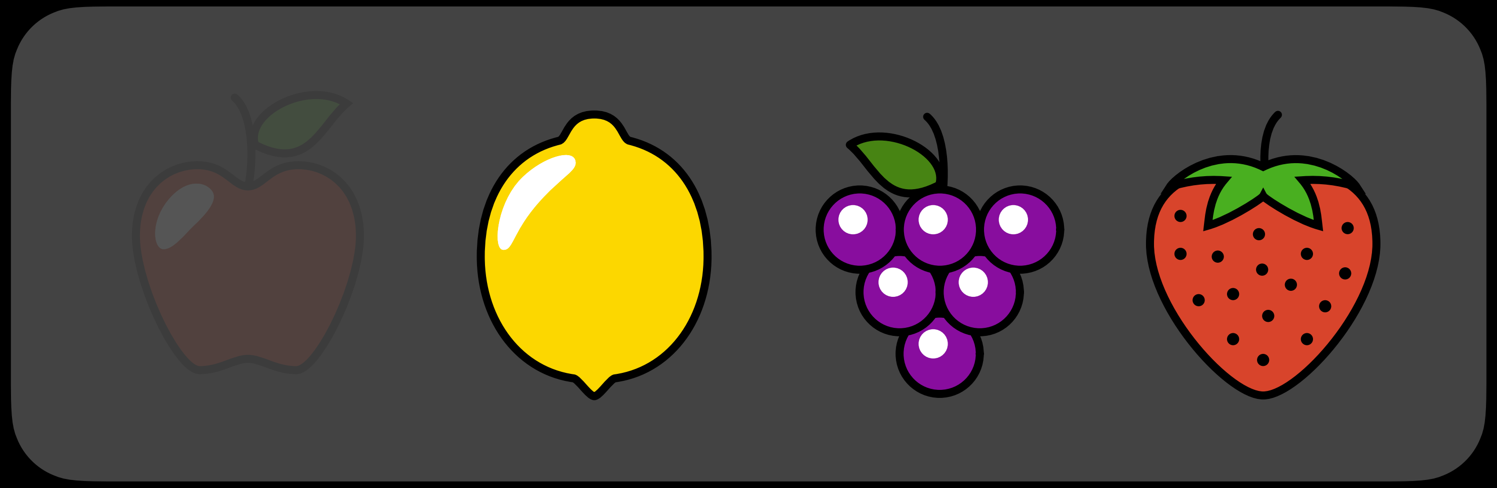




“The fundamental problem of discrete choice”

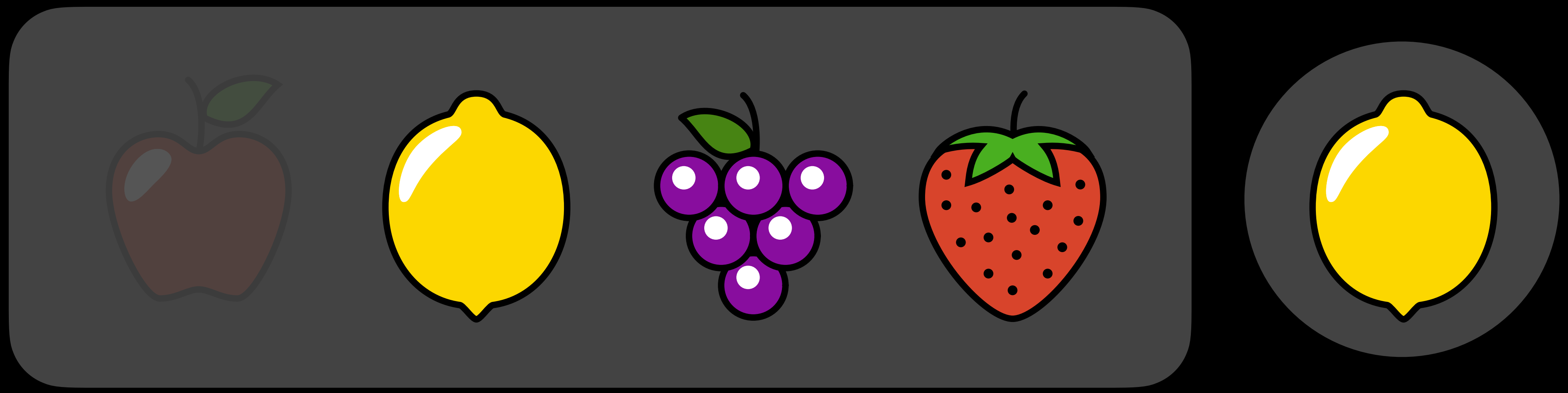
“The fundamental problem of discrete choice”

choice set



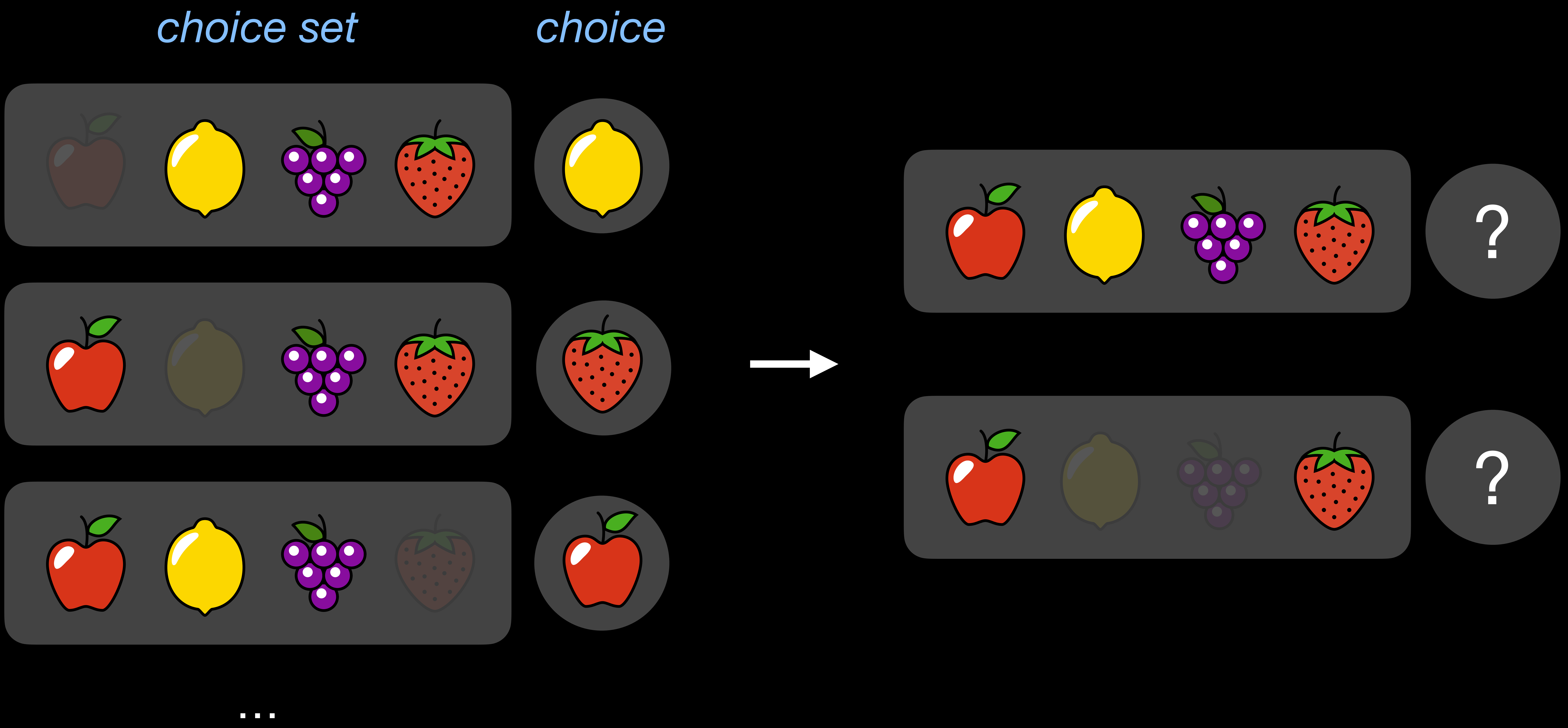
“The fundamental problem of discrete choice”

choice set *choice*



“The fundamental problem of discrete choice”

“The fundamental problem of discrete choice”



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

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

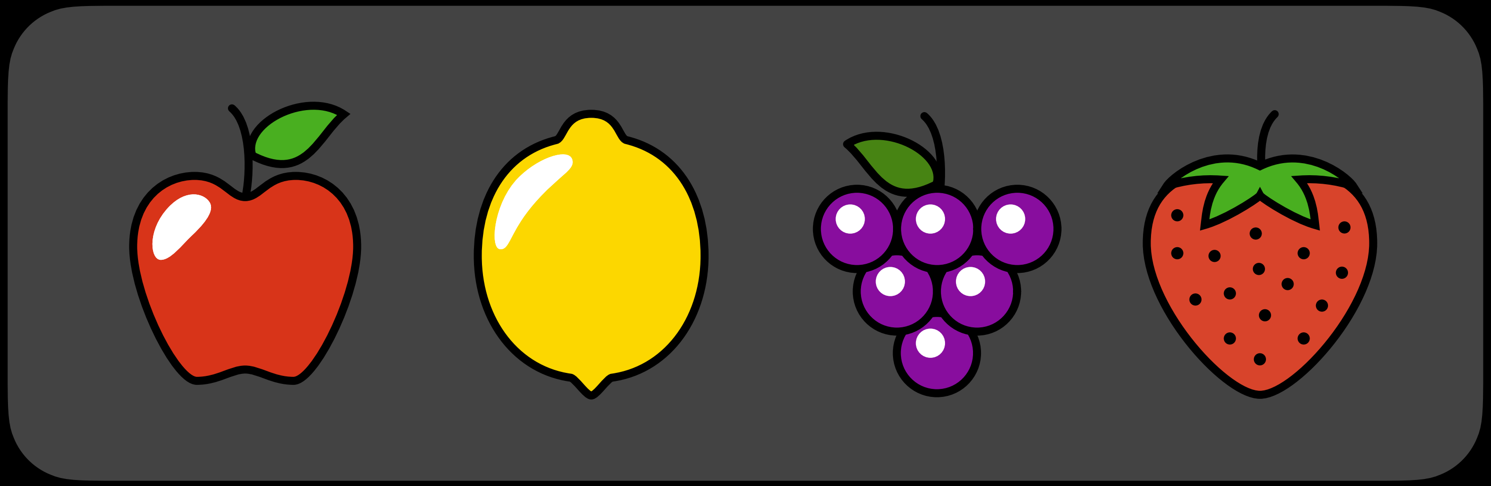
The classic model: *multinomial logit (MNL)*

(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



u_i	1	-1	0	2
$\Pr(i \mid C)$	.24	.03	.09	.64

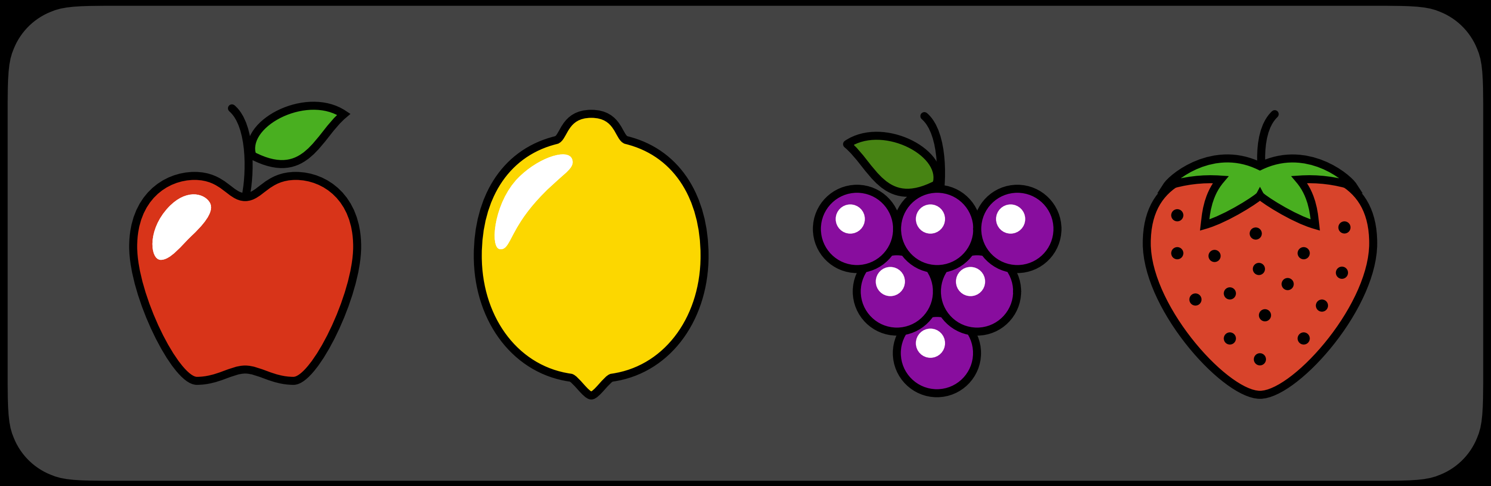
The classic model: *multinomial logit (MNL)*

(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



u_i	1	-1	0	2
$\Pr(i \mid C)$	.24	.03	.09	.64

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

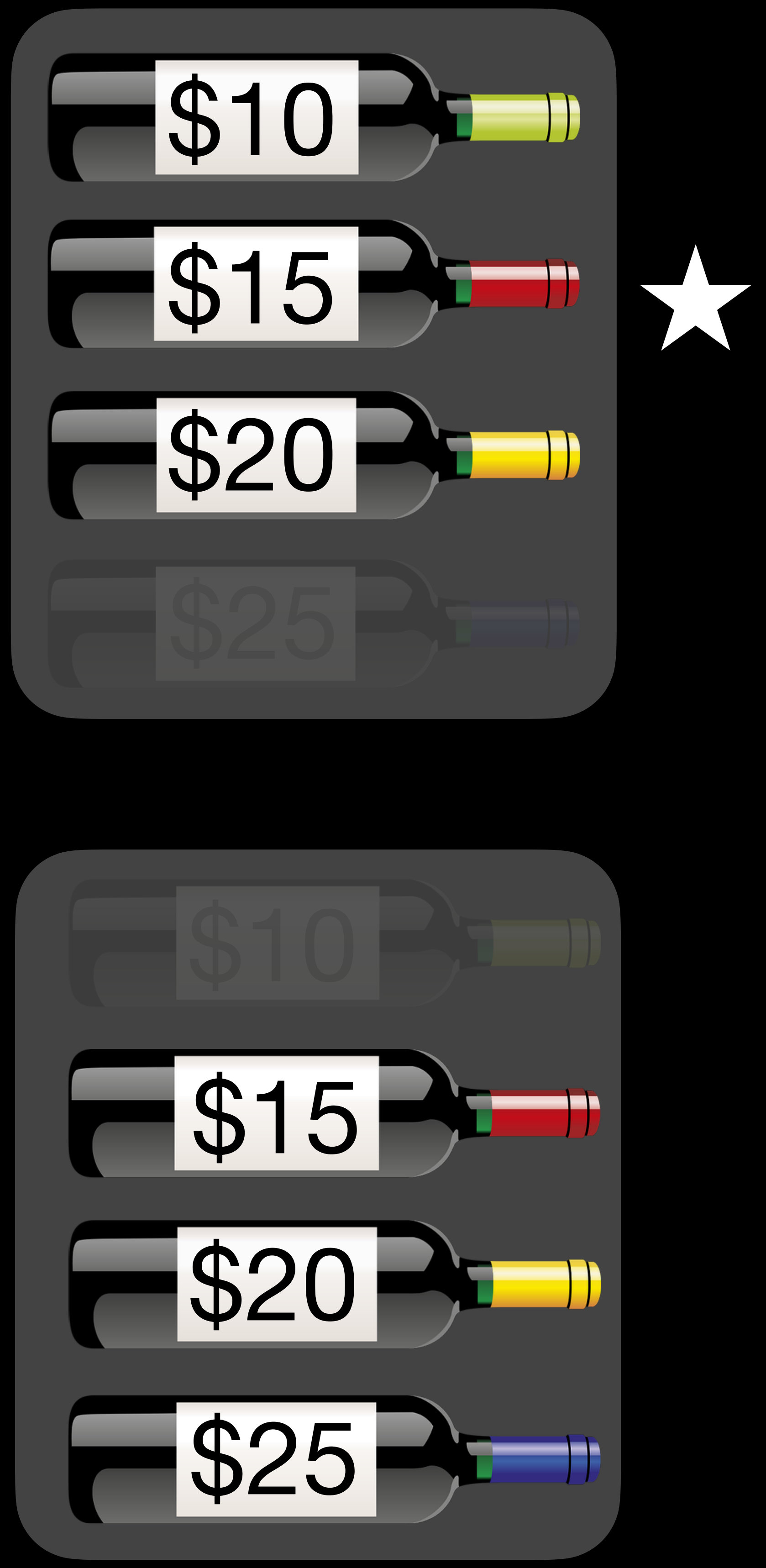
(Simonson, 1989)

Problem for MNL: *context effects*

The choice set influences preferences.

Compromise

(Simonson, 1989)

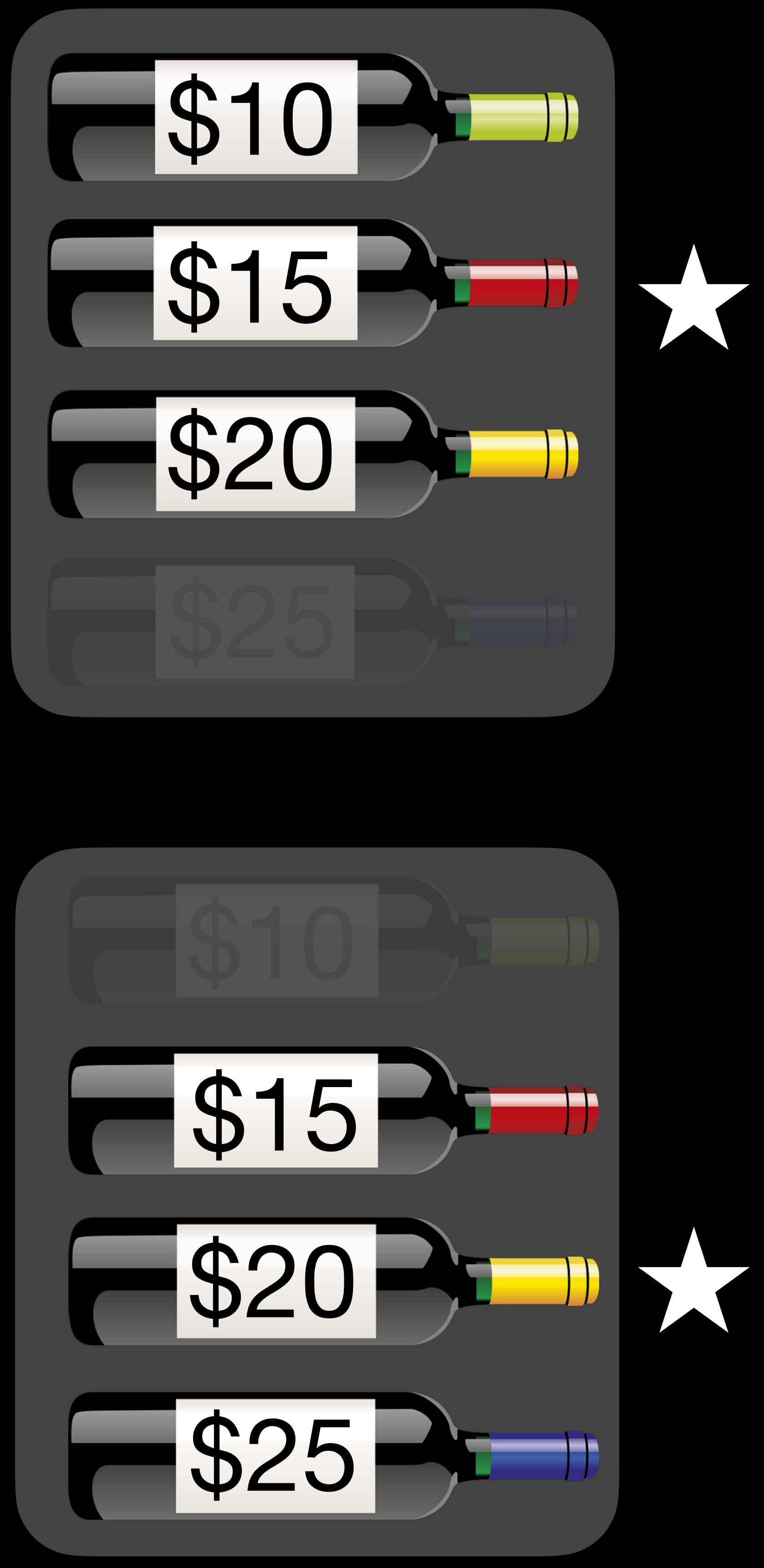


Problem for MNL: *context effects*

The choice set influences preferences.

Compromise

(Simonson, 1989)

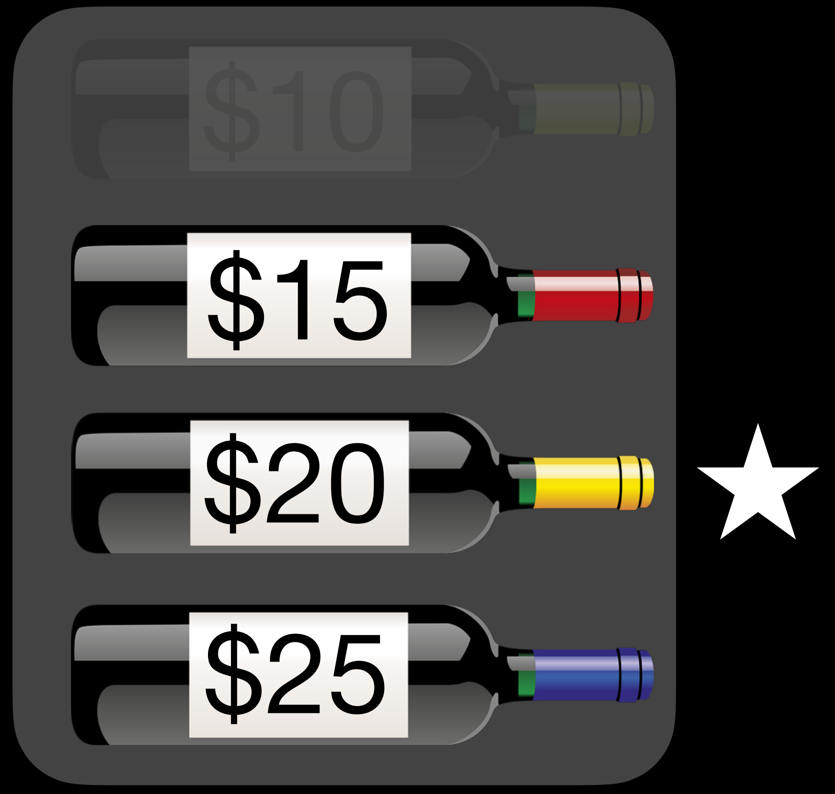


Problem for MNL: *context effects*

The choice set influences preferences.

Compromise

(Simonson, 1989)



Similarity

(Tversky, 1972)

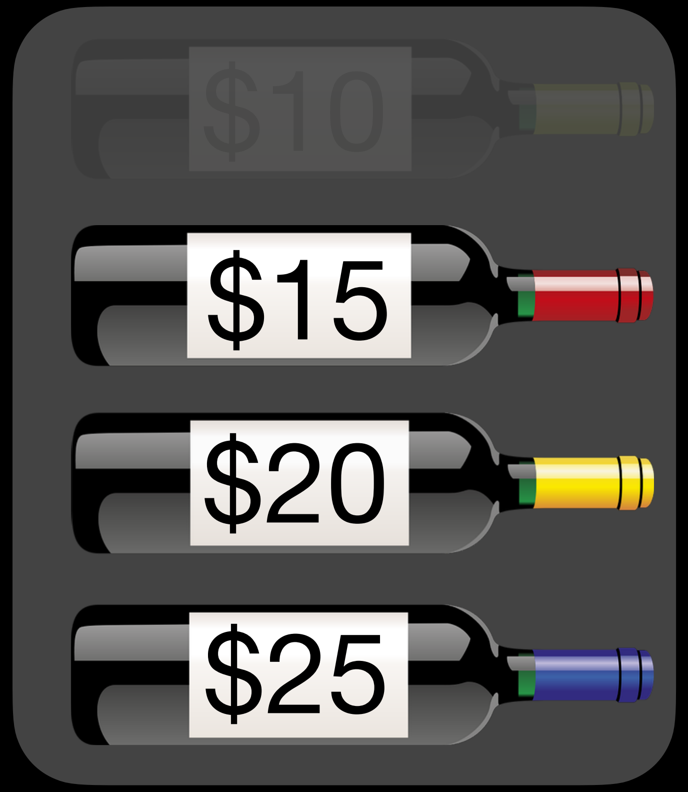
Problem for MNL: *context effects*

The choice set influences preferences.

Compromise

(Simonson, 1989)

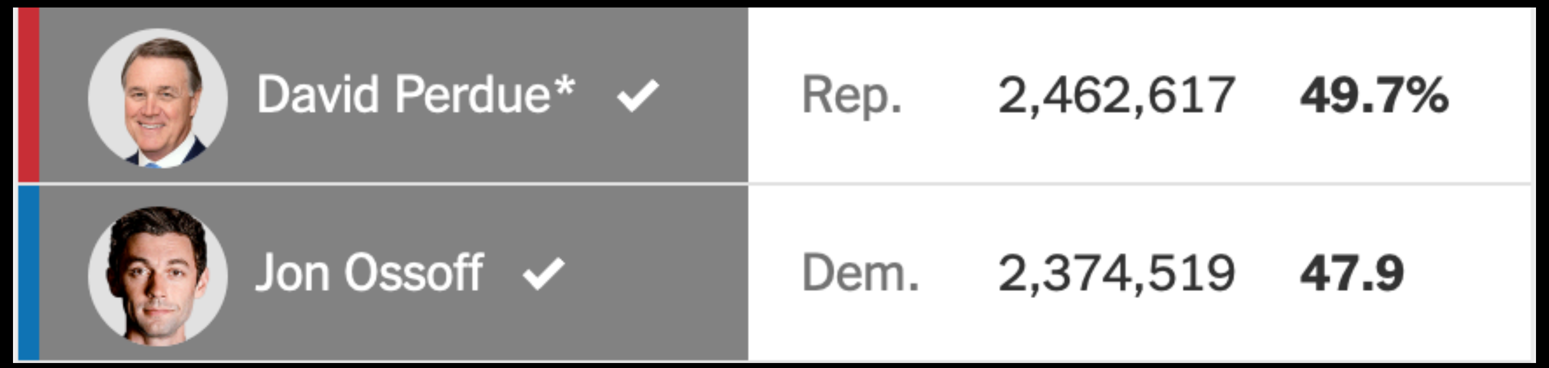
★



★

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)

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**

Problem for MNL: *context effects*

The choice set influences preferences.

Compromise
(Simonson, 1989)

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:

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

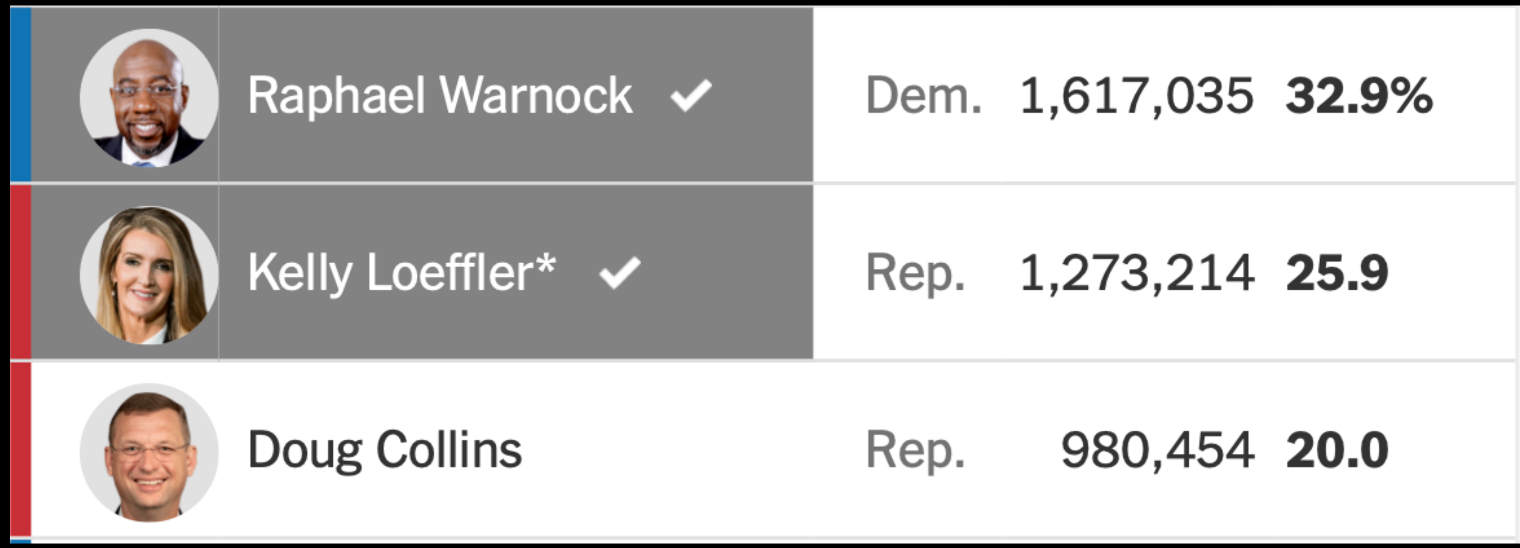
Assumes no higher-order effects
(i.e., context effects decompose additively into effects of items)

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:

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



Raphael Warnock ✔	Dem.	1,617,035	**32.9%**
Kelly Loeffler* ✔	Rep.	1,273,214	**25.9**
Doug Collins	Rep.	980,454	**20.0**

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

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

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: *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: *conditional logit*

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

MNL:

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

$\longrightarrow$

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:

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

$\longrightarrow$

Conditional logit:

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

Preference coefficient θ_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: $u_i = \theta^T x_i$ ⟶ *Contextual utility*: $u_{i,C} = [\theta + F(C)]^T x_i$

Incorporating *feature context effects* into conditional logit

Conditional logit utility: $u_i = \theta^T x_i$ $\longrightarrow$ *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: $u_i = \theta^T x_i$ $\longrightarrow$ *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: $u_i = \theta^T x_i$ $\longrightarrow$ *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: $u_i = \theta^T x_i$ $\longrightarrow$ *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*: $f(x_j) = A x_j$ for some matrix $A \in \mathbb{R}^{d \times d}$

Incorporating *feature context effects* into conditional logit

Conditional logit utility: $u_i = \theta^T x_i$ ⟶ *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 $A \in \mathbb{R}^{d \times d}$

→ $u_{i,C} = (\theta + Ax_C)^T x_i$ ($x_C = \frac{1}{|C|} \sum_{j \in C} x_j$ 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

→ θ: *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

→ θ: *base preference coefficients*

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

The *Linear Context Logit (LCL)*

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

→ convex negative log-likelihood

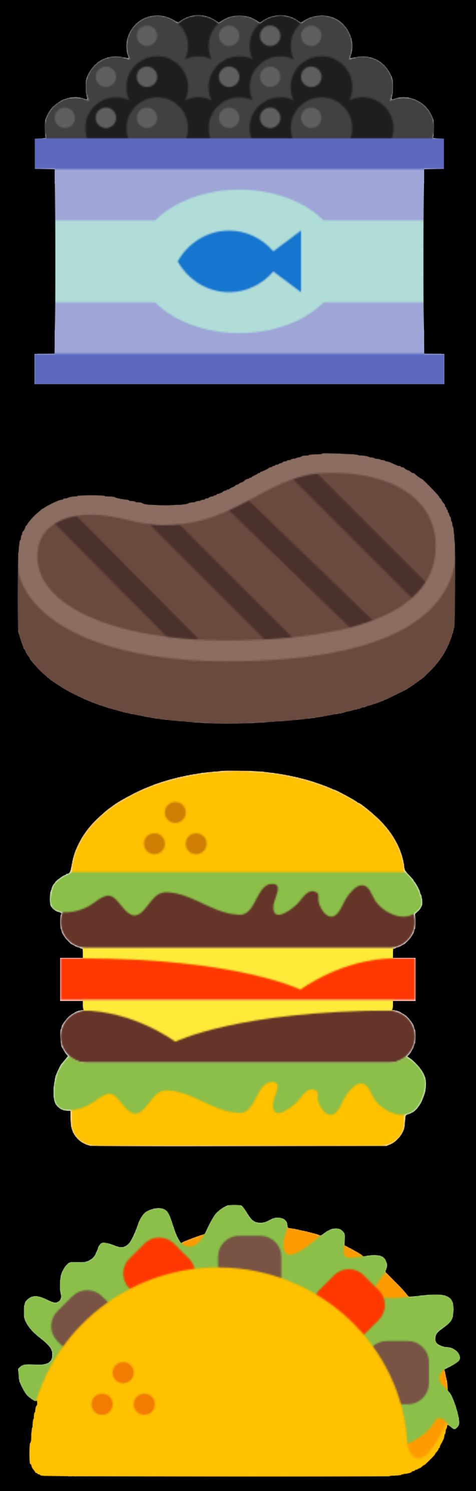
→ θ: *base preference coefficients*

→ $A_{pq} > 0$: 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

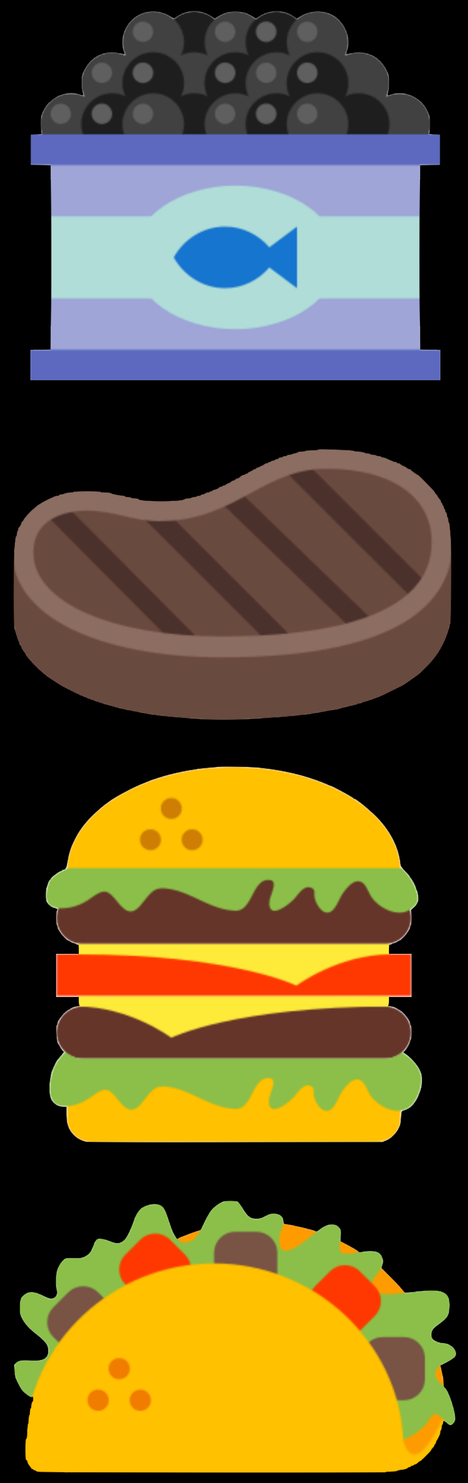
LCL example: restaurant selection

LCL example: restaurant selection



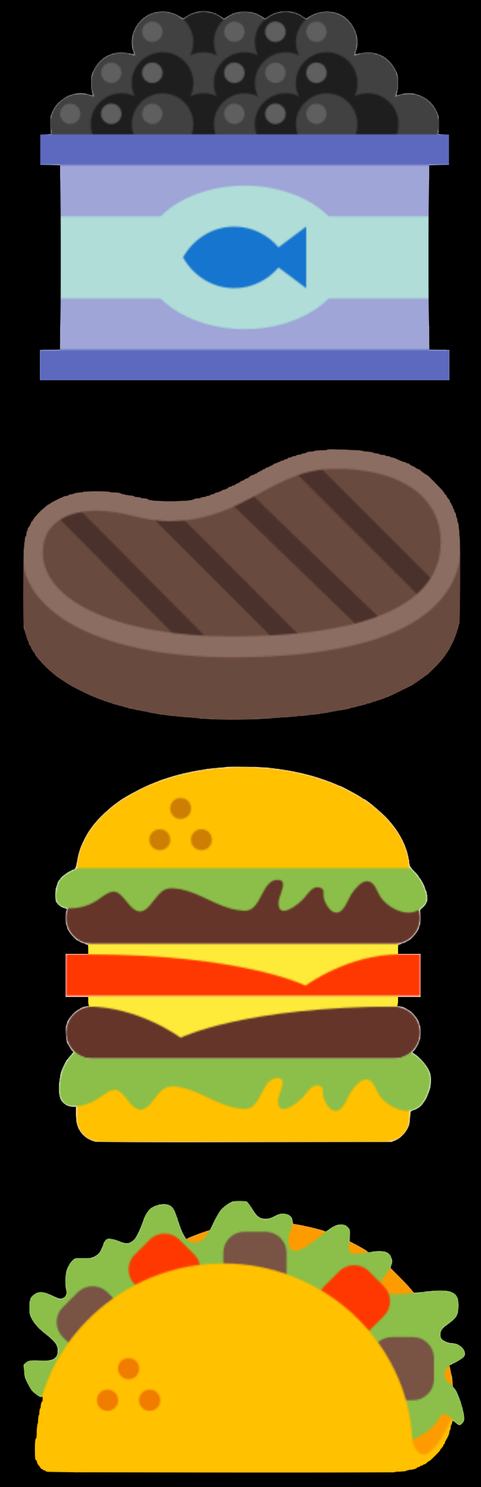
Icons by icons8.com

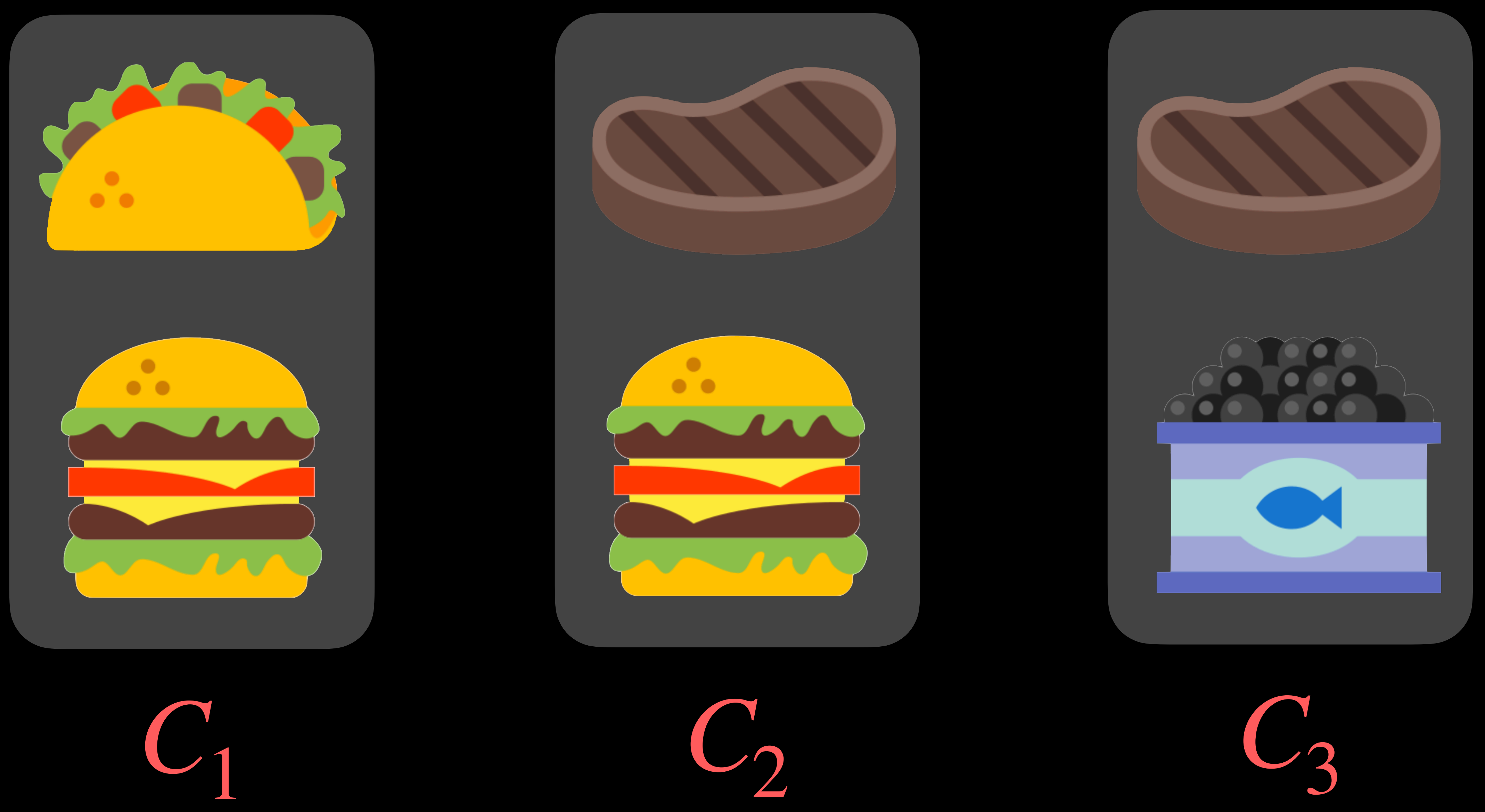
LCL example: restaurant selection



item features:
- price
- service speed
- wine selection

LCL example: restaurant selection





item features:
- price
- service speed
- wine selection

LCL example: restaurant selection

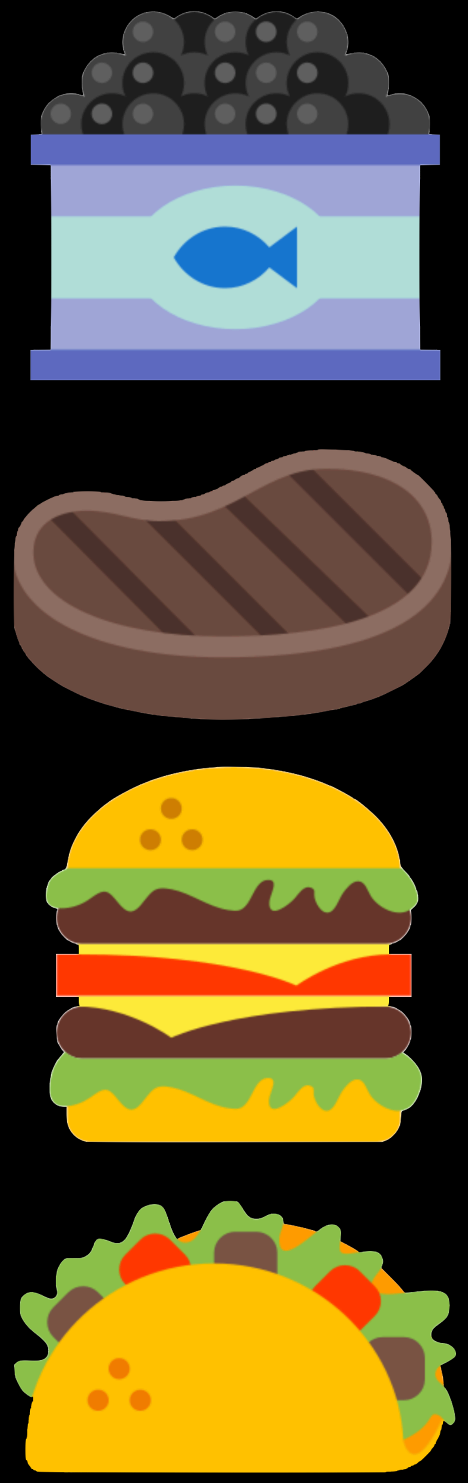
item features:
- price
- service speed
- wine selection

C_1 C_2 C_3

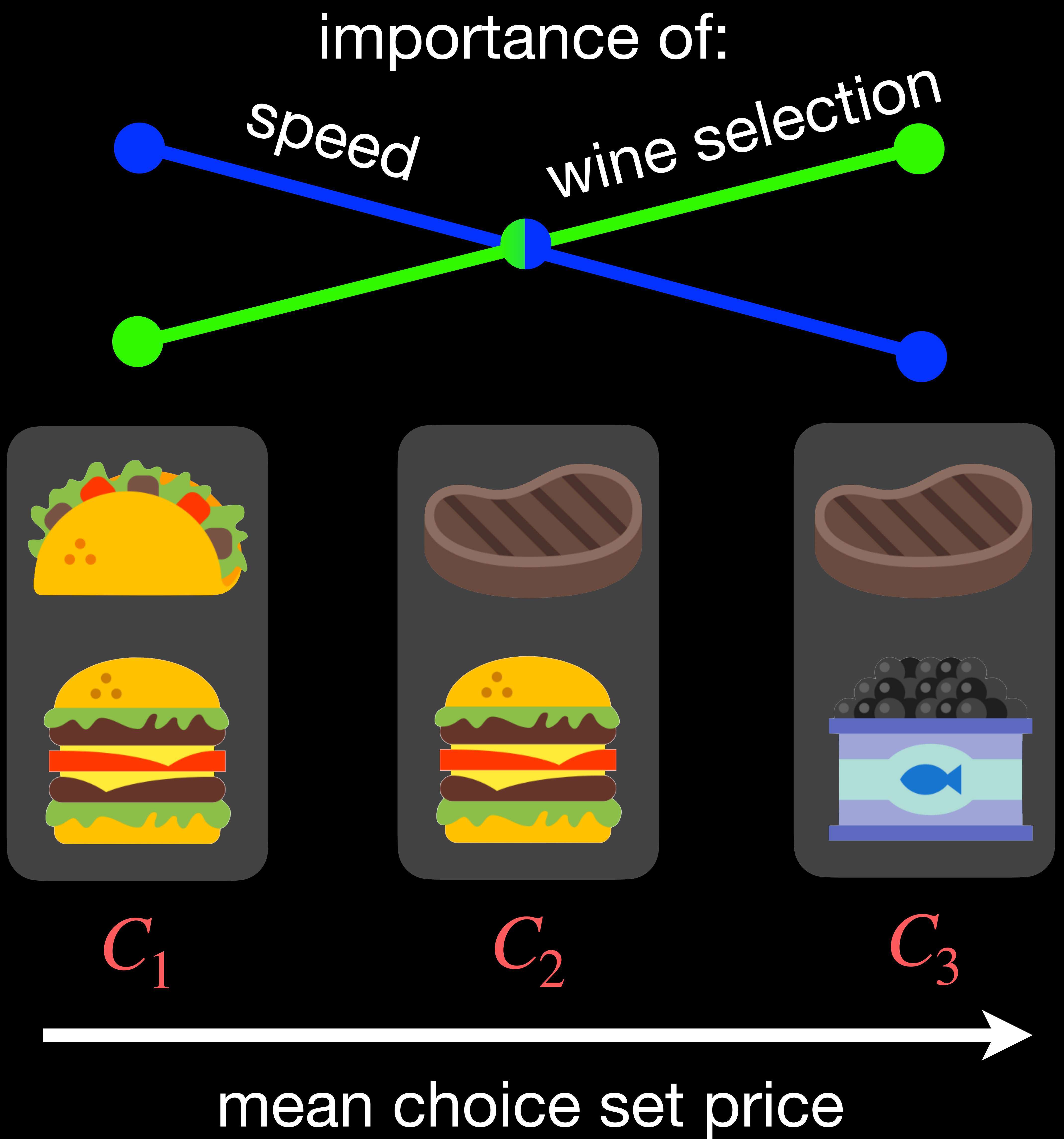
mean choice set price

Icons by icons8.com

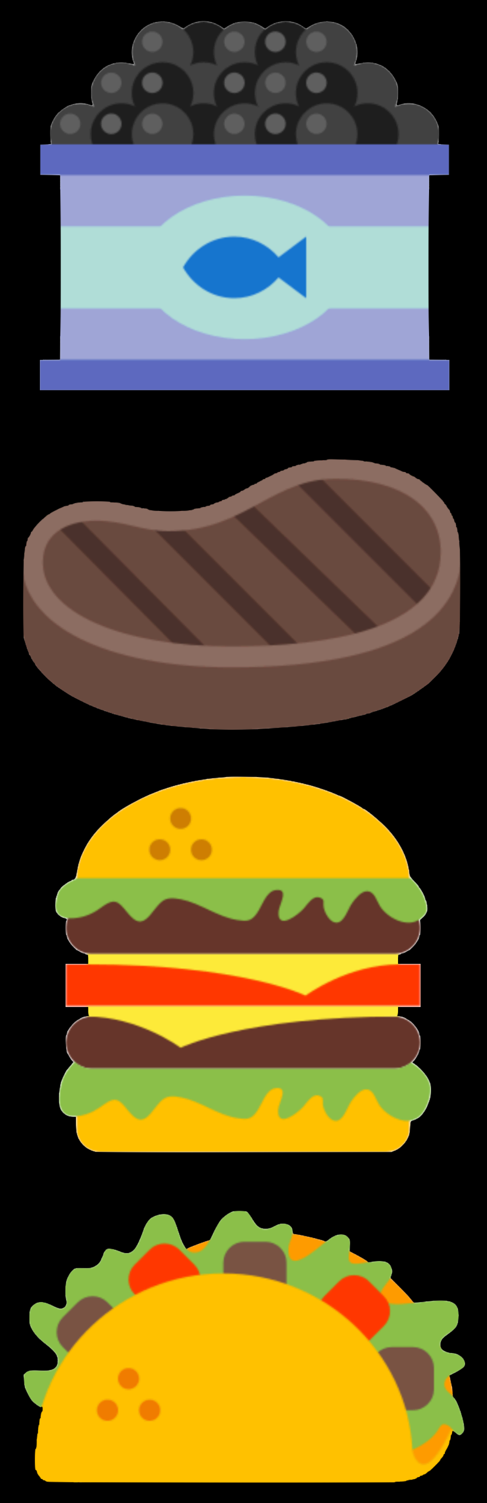
LCL example: restaurant selection



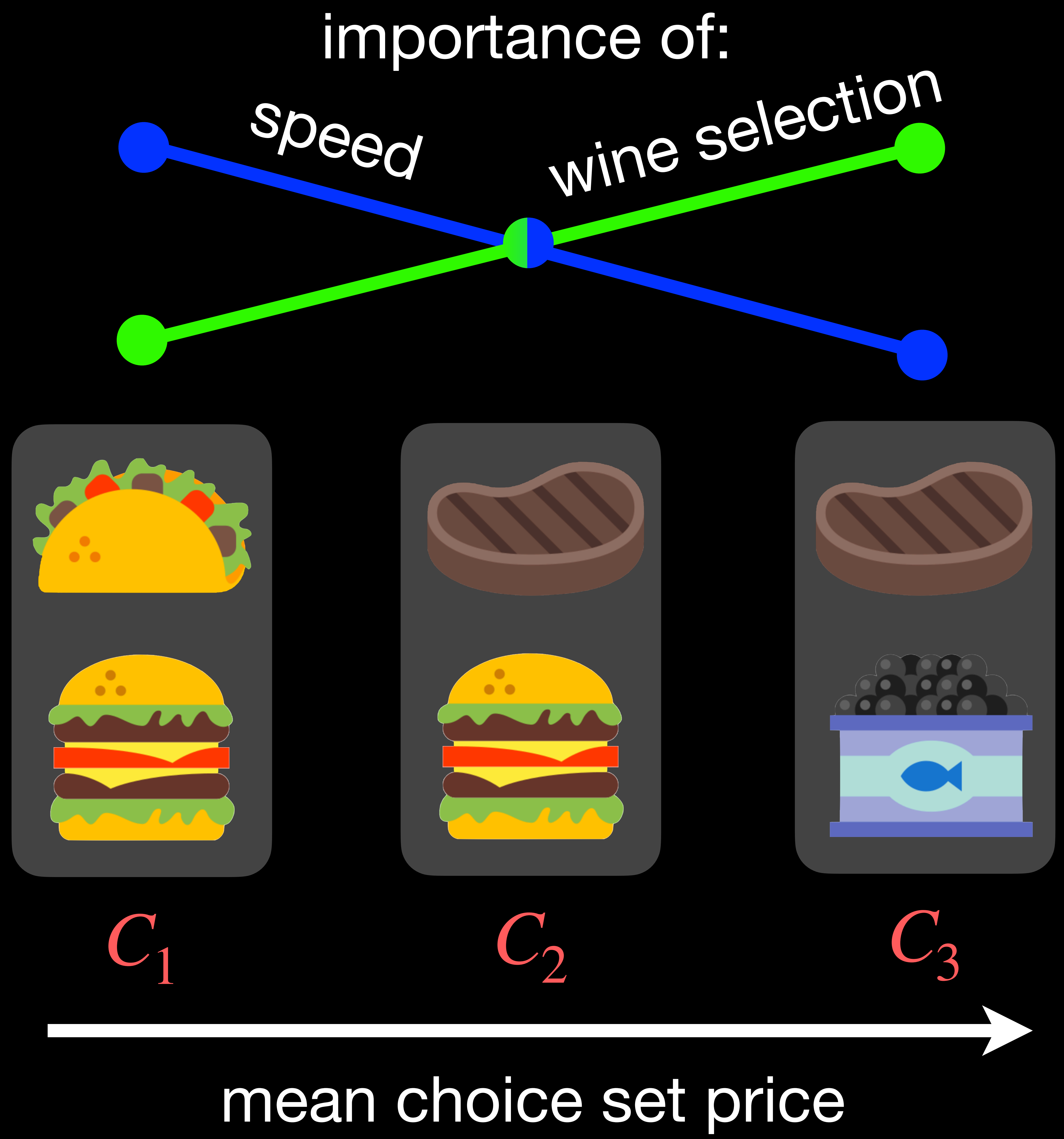
item features:
- price
- service speed
- wine selection



LCL example: restaurant selection



item features:
- price
- service speed
- wine selection



$$A = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

LCL identifiability, fully characterized

LCL identifiability, fully characterized

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

LCL identifiability, fully characterized

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

→ important for inference and interpretation

LCL identifiability, fully characterized

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

→ important for inference and interpretation

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

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

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

LCL identifiability, fully characterized

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

→ important for inference and interpretation

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

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

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

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

Dataset	Choices	Features	Largest Choice Set
DISTRICT	5376	27	2
DISTRICT-SMART	5376	6	2
SUSHI	5000	6	10
EXPEDIA	276593	5	38
CAR-A	2675	4	2
CAR-B	2206	5	2
CAR-ALT	4654	21	6

Choice datasets

favorite sushi types

Dataset	Choices	Features	Largest Choice Set
DISTRICT	5376	27	2
DISTRICT-SMART	5376	6	2
SUSHI	5000	6	10
EXPEDIA	276593	5	38
CAR-A	2675	4	2
CAR-B	2206	5	2
CAR-ALT	4654	21	6

Choice datasets

favorite sushi types

hotel bookings

Dataset	Choices	Features	Largest Choice Set
DISTRICT	5376	27	2
DISTRICT-SMART	5376	6	2
SUSHI	5000	6	10
EXPEDIA	276593	5	38
CAR-A	2675	4	2
CAR-B	2206	5	2
CAR-ALT	4654	21	6

LCL improves model fit

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

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

*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

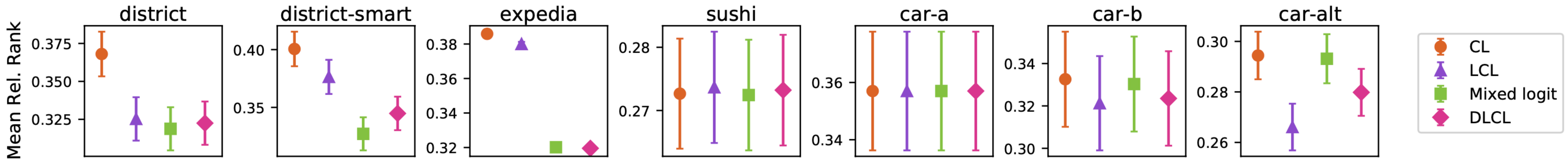


Figure 2: Mean relative rank of predictions on held-out test data (lower is better). Error bars show standard error of the mean.

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 4: Five largest context effects in SUSHI.

Effect (q on p)	A_{pq} (std. err.)	p-value
popularity on *popularity*	−0.28 (0.15)	0.066
availability on *is maki*	0.24 (0.14)	0.087
oiliness on *oiliness*	−0.20 (0.08)	0.0089
popularity on *availability*	0.19 (0.14)	0.16
availability on *oiliness*	−0.18 (0.10)	0.064

LCL can test individual effects for significance

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

Table 4: Five largest context effects in SUSHI.

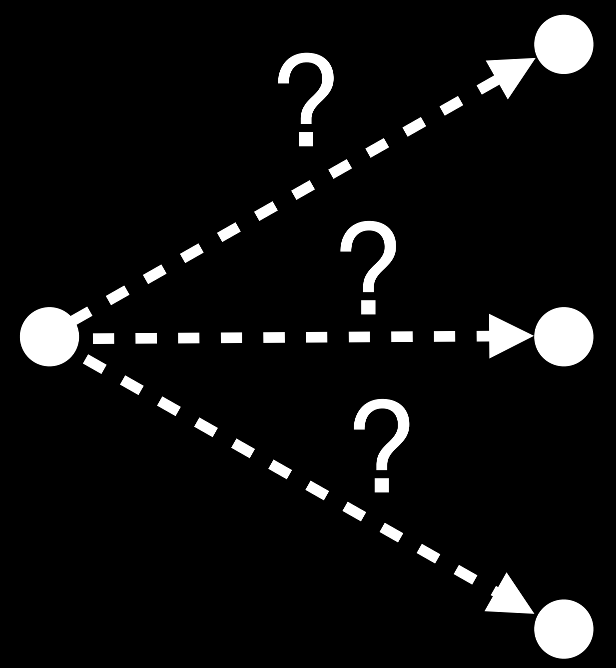
Effect (q on p)	A_{pq} (std. err.)	p-value
popularity on *popularity*	−0.28 (0.15)	0.066
availability on *is maki*	0.24 (0.14)	0.087
oiliness on *oiliness*	−0.20 (0.08)	0.0089
popularity on *availability*	0.19 (0.14)	0.16
availability on *oiliness*	−0.18 (0.10)	0.064

Table 5: Five largest context effects in EXPEDIA.

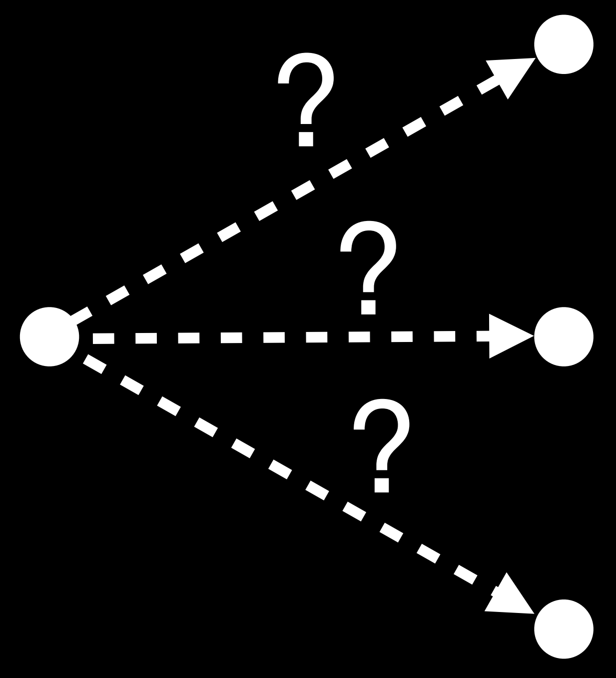
Effect (q on p)	A_{pq} (std. err.)	p-value
location score on *price*	−0.47 (0.05)	$< 10^{-16}$
on promotion on *price*	0.27 (0.03)	$< 10^{-16}$
review score on *price*	−0.19 (0.03)	1.4×10^{-9}
star rating on *price*	0.15 (0.04)	6.7×10^{-5}
price on *star rating*	0.10 (0.00)	$< 10^{-16}$

Social network application

What factors drive edge formation?

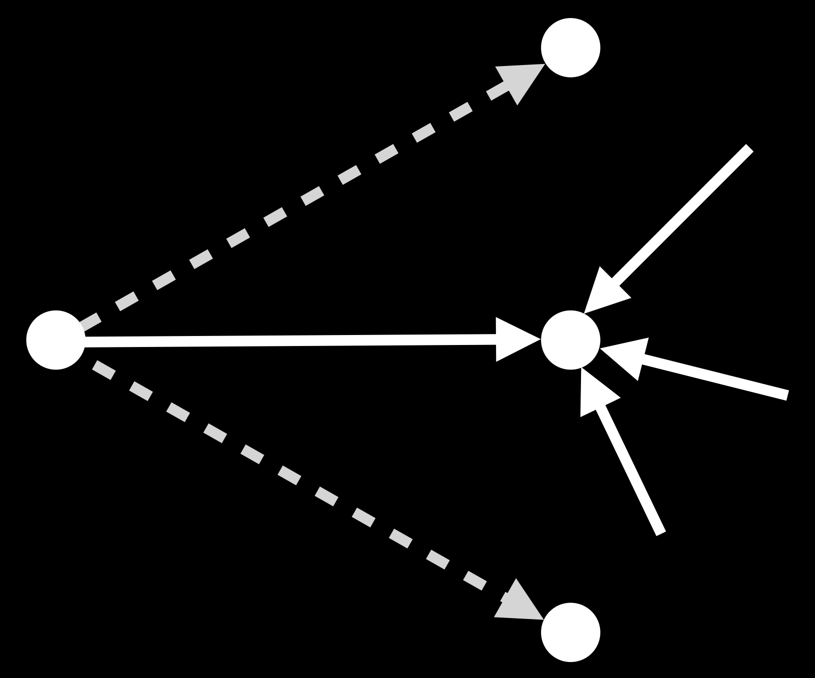


What factors drive edge formation?

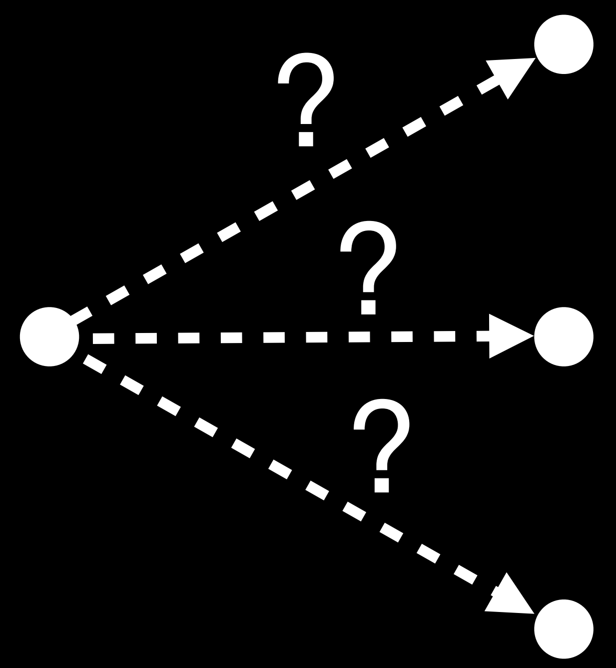


Preferential attachment

(Barabási & Albert, *Science* 1999)

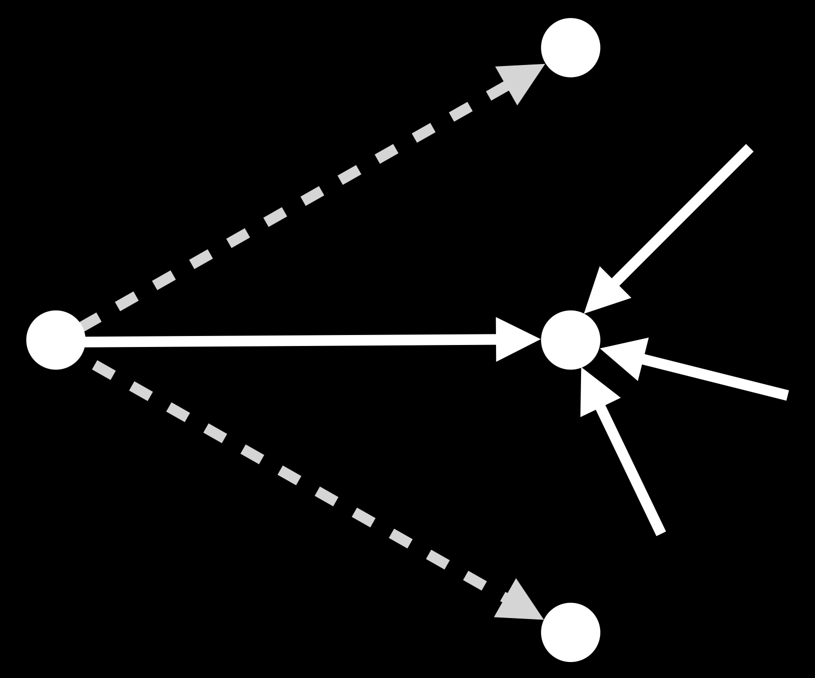


What factors drive edge formation?



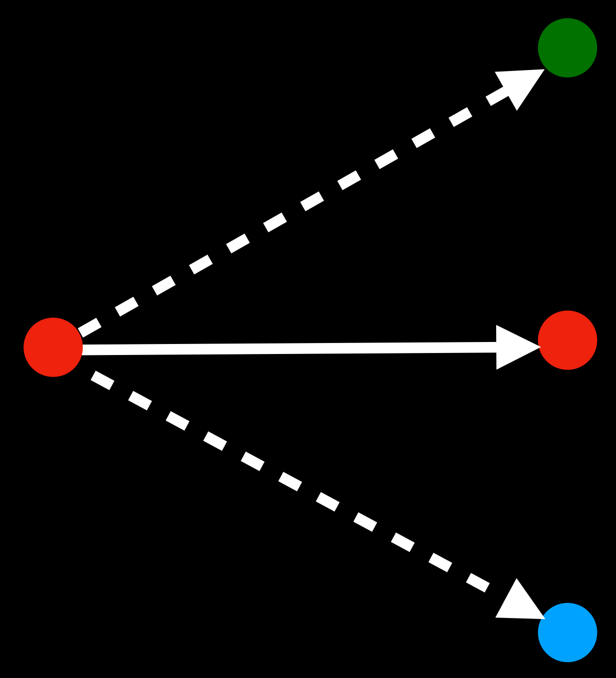
Preferential attachment

(Barabási & Albert, *Science* 1999)

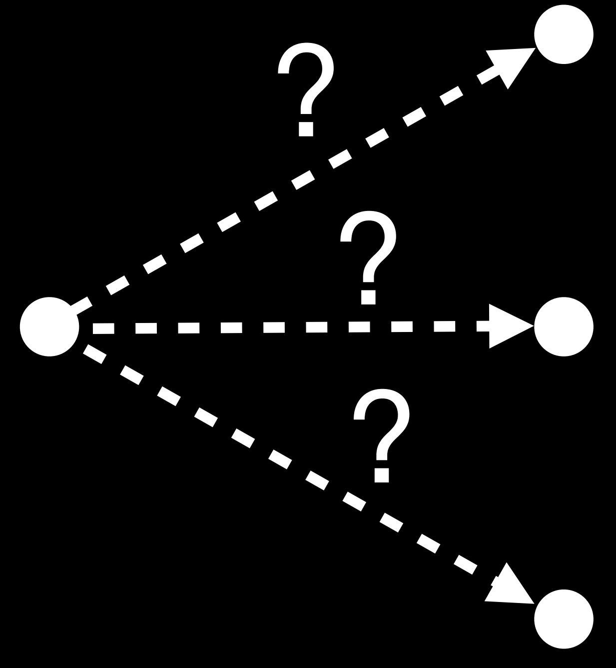


Homophily

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

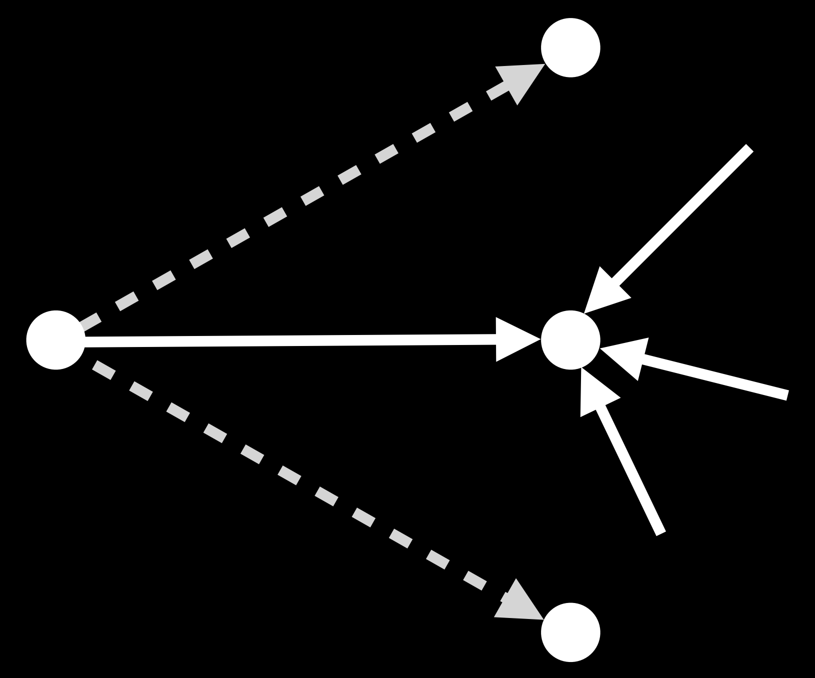


What factors drive edge formation?



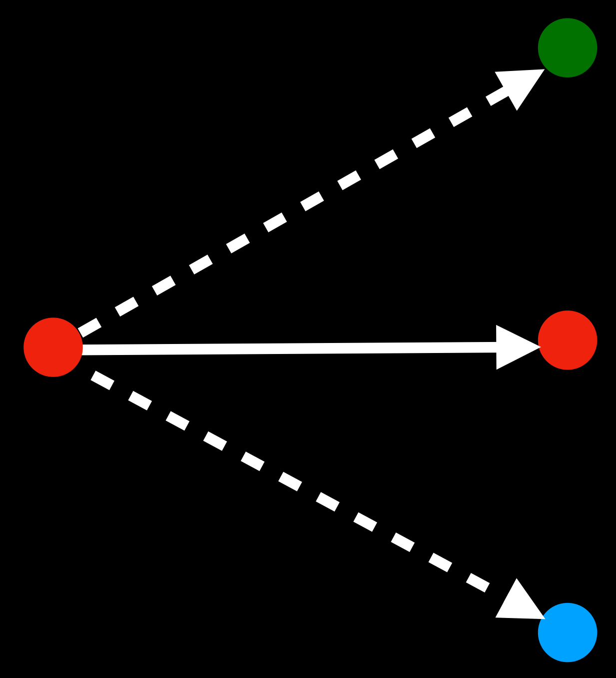
Preferential attachment

(Barabási & Albert, *Science* 1999)



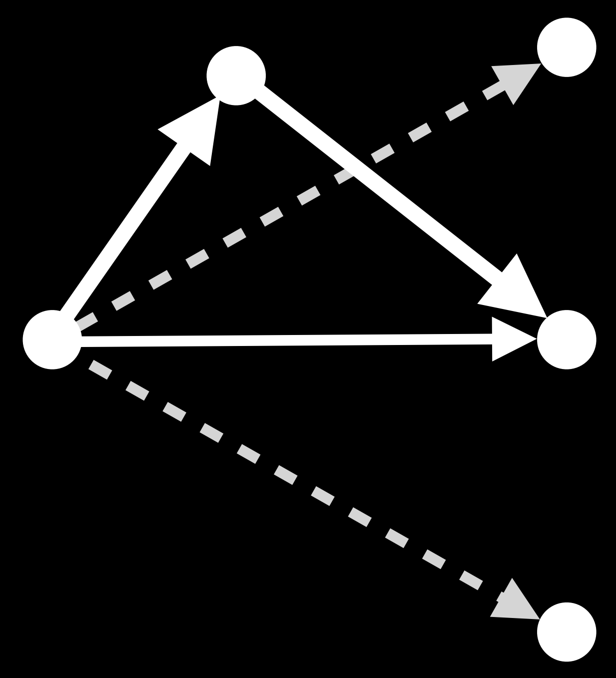
Homophily

(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)



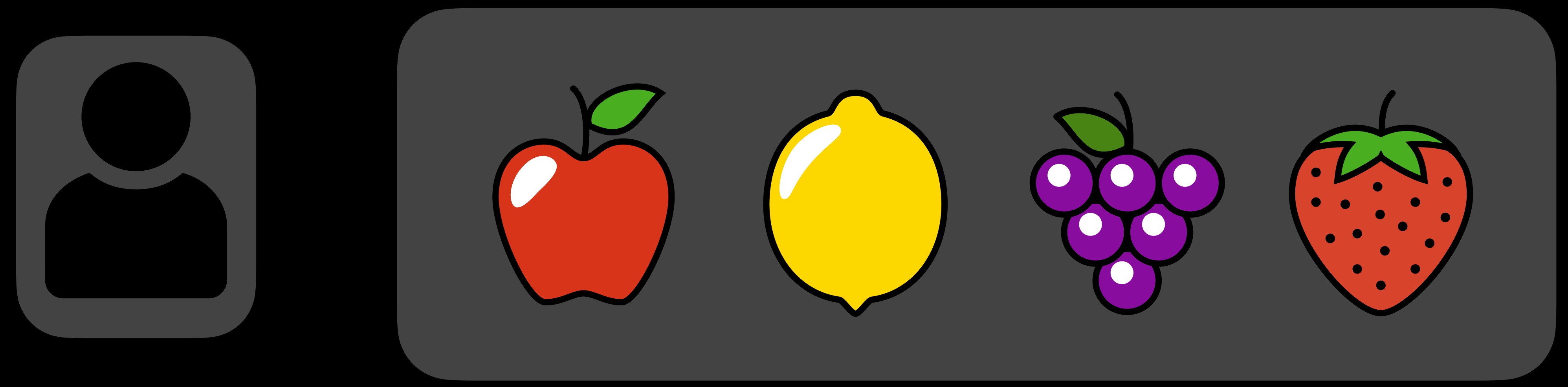
“Choosing to grow a graph”

(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:

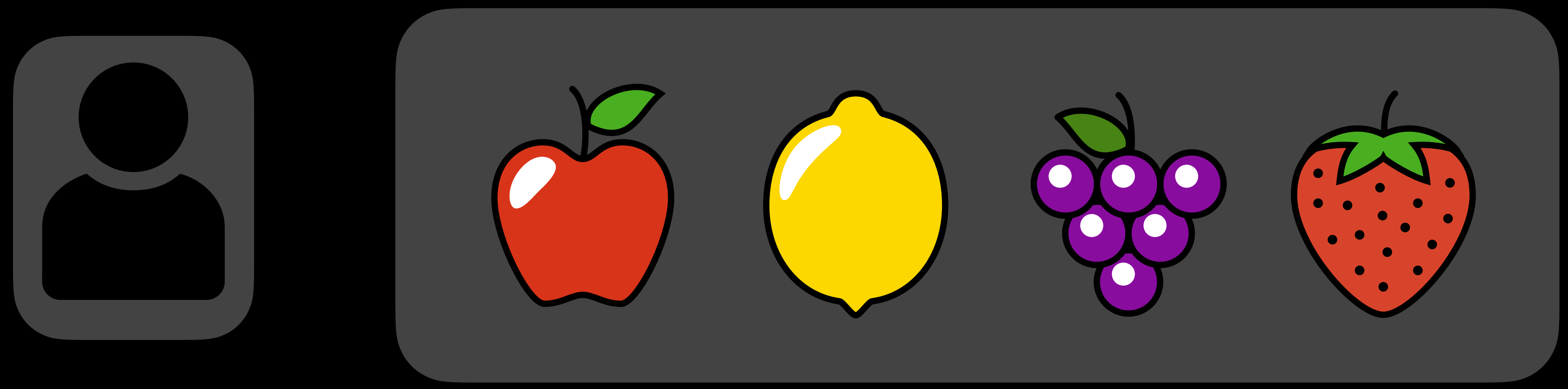


chooser *choice set*

“Choosing to grow a graph”

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

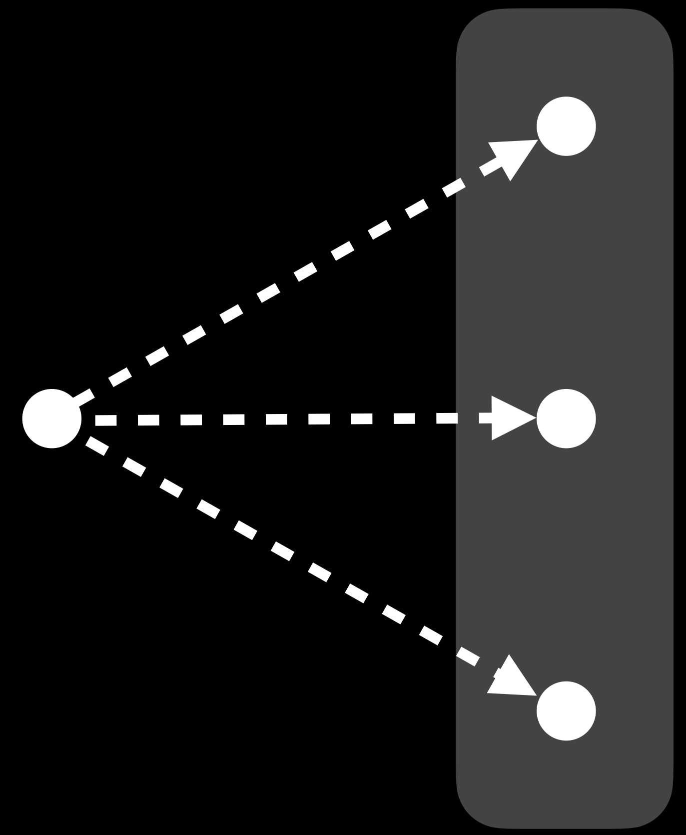
so far:



chooser

choice set

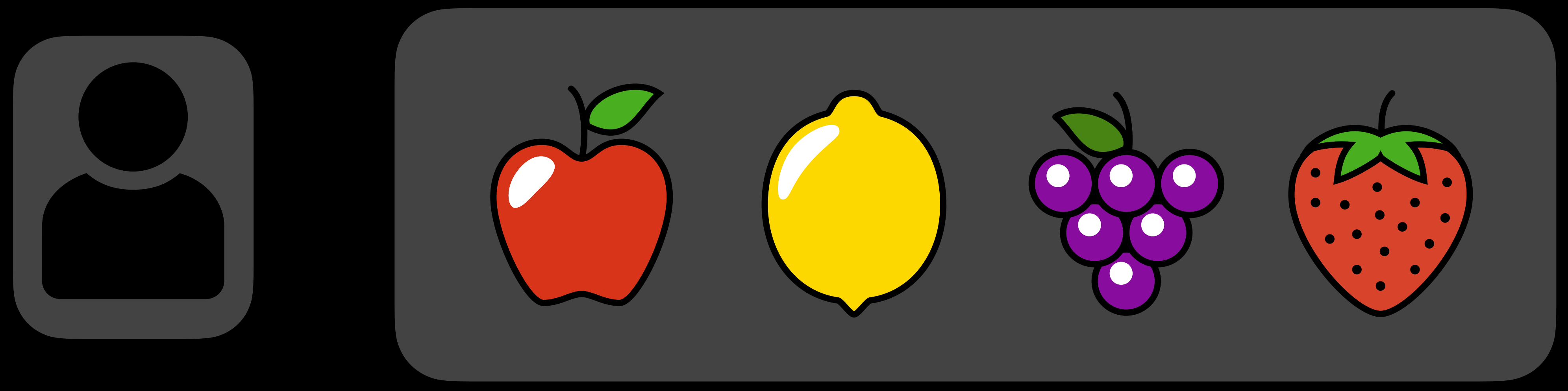
in network growth:



“Choosing to grow a graph”

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

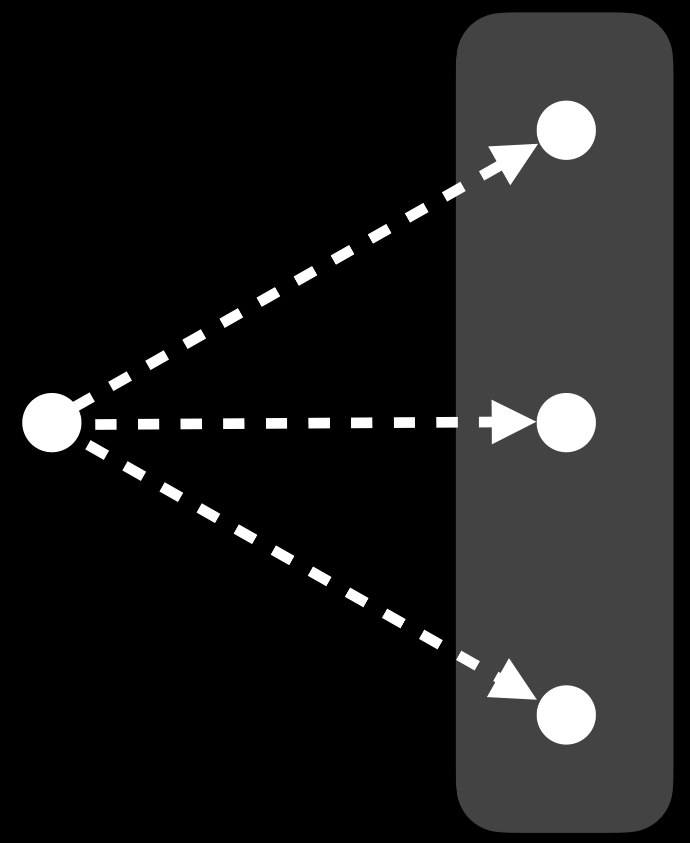
so far:



chooser *choice set*

in network growth:

chooser



choice set

Key usage

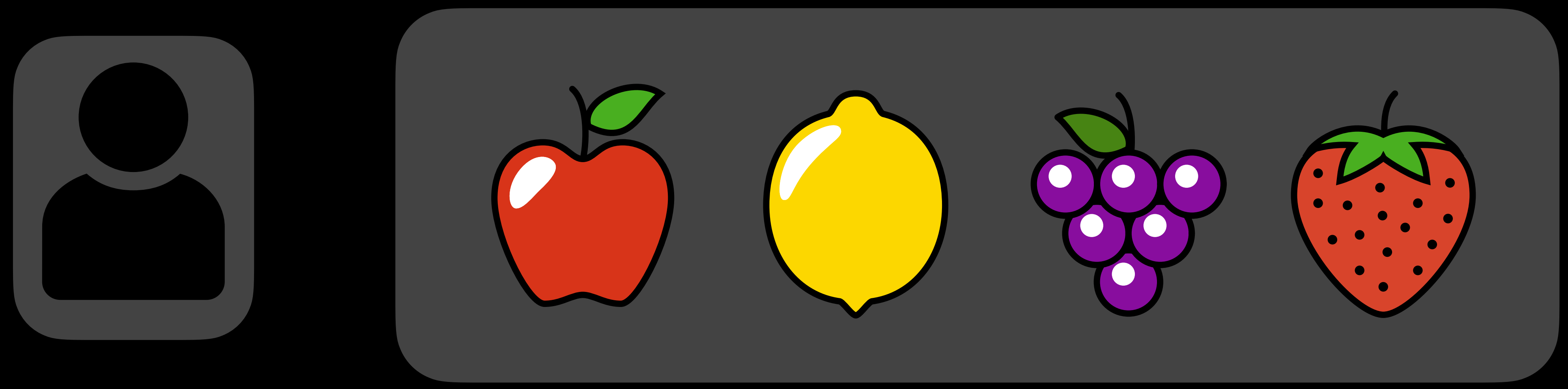
Timestamped edges
→ meaningful choice sets

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)

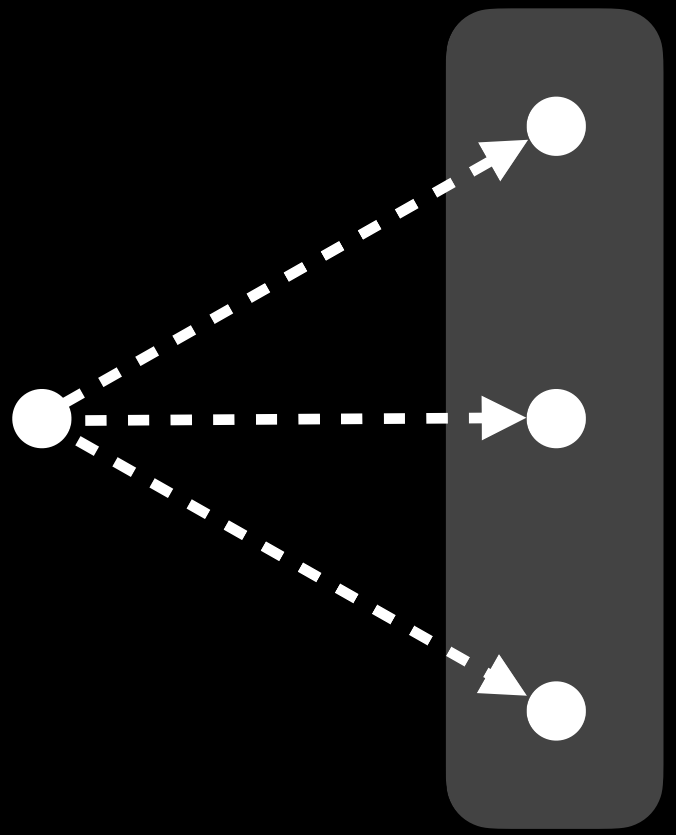
so far:



chooser *choice set*

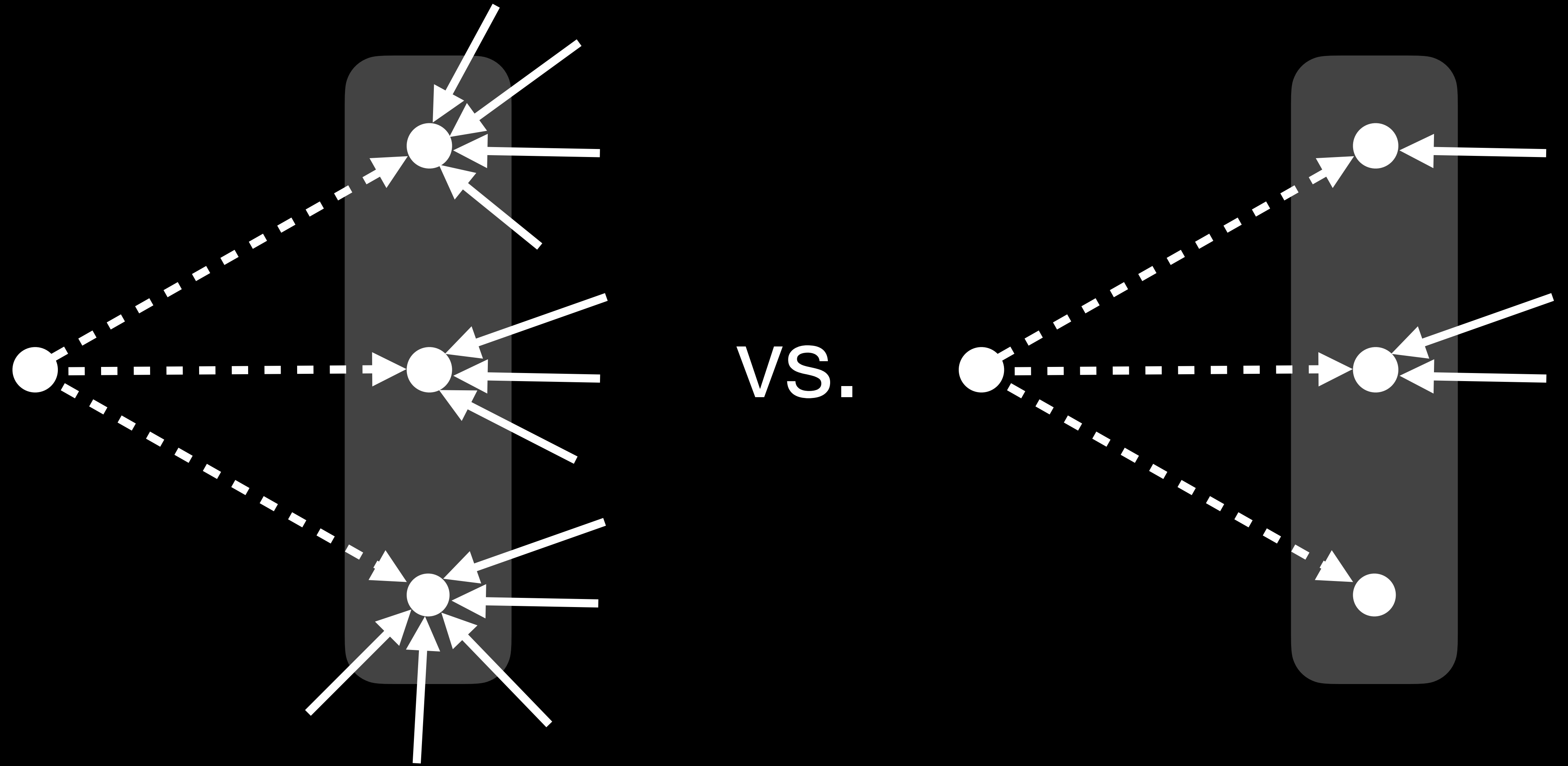
in network growth:

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

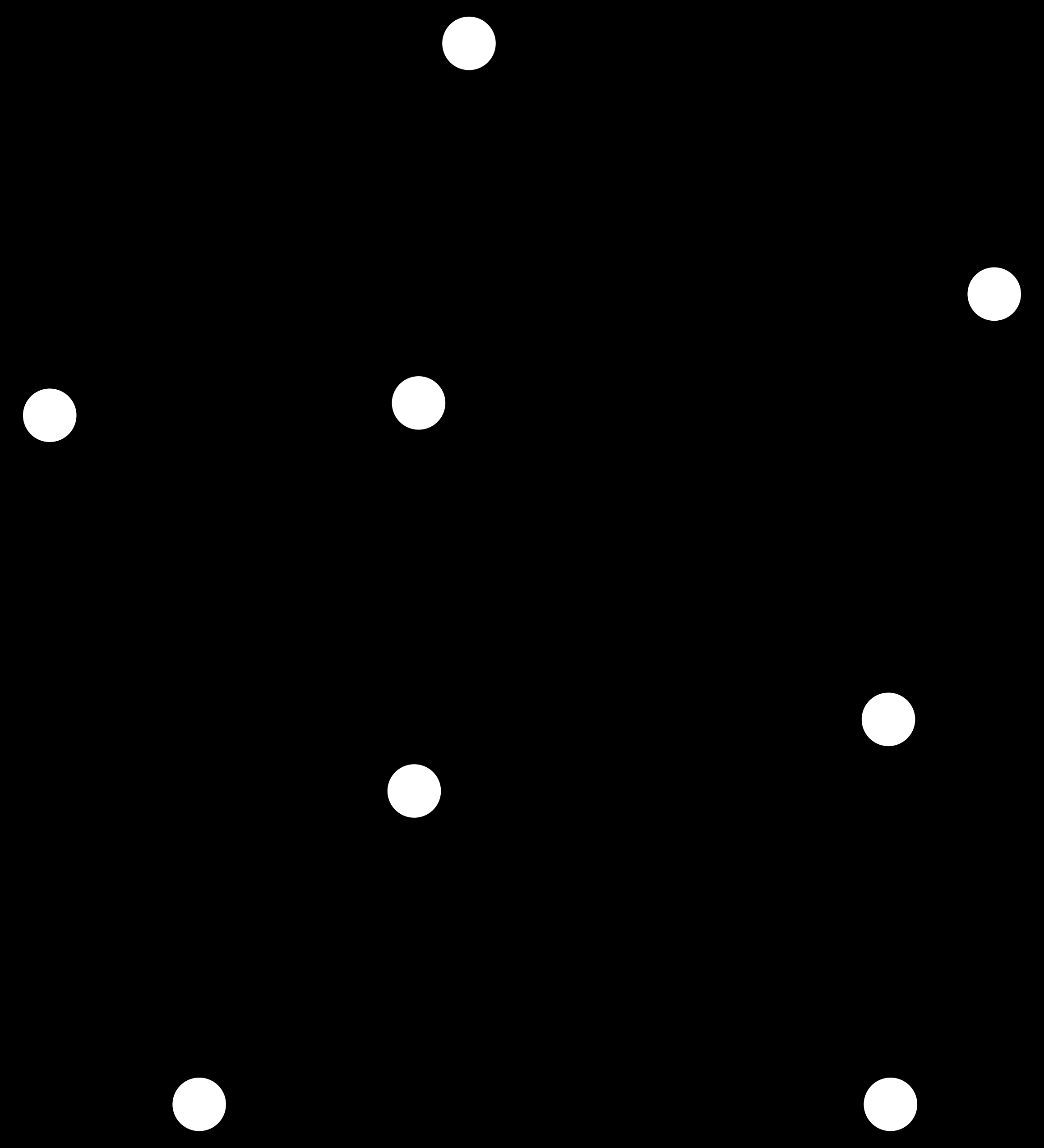
Choosing to close triangles

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

Our data
Timestamped edges
(including repeats)

Choosing to close triangles

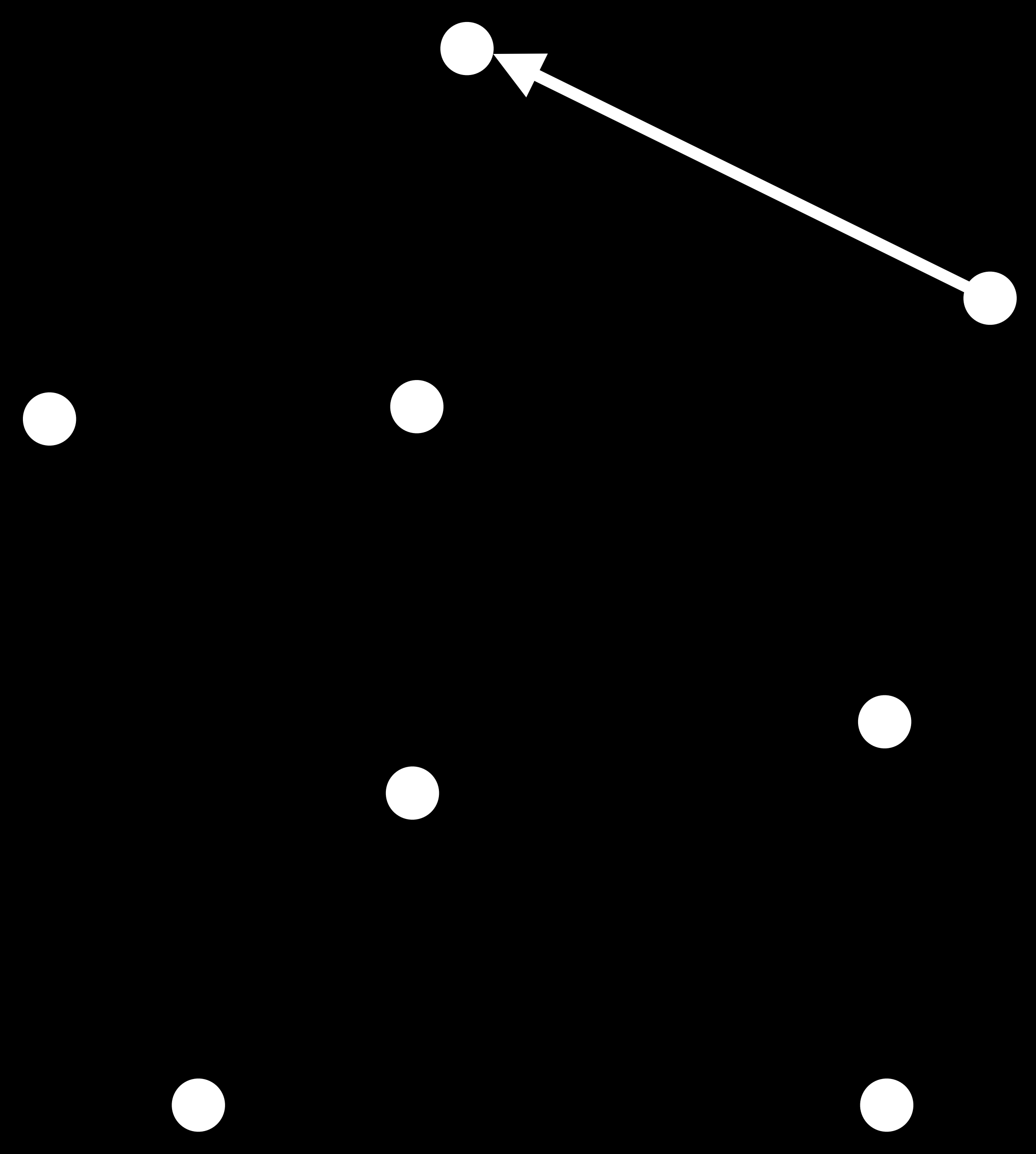
Triadic closure offers small choice sets
→ tractable inference
→ varied choice sets



Our data
Timestamped edges
(including repeats)

Choosing to close triangles

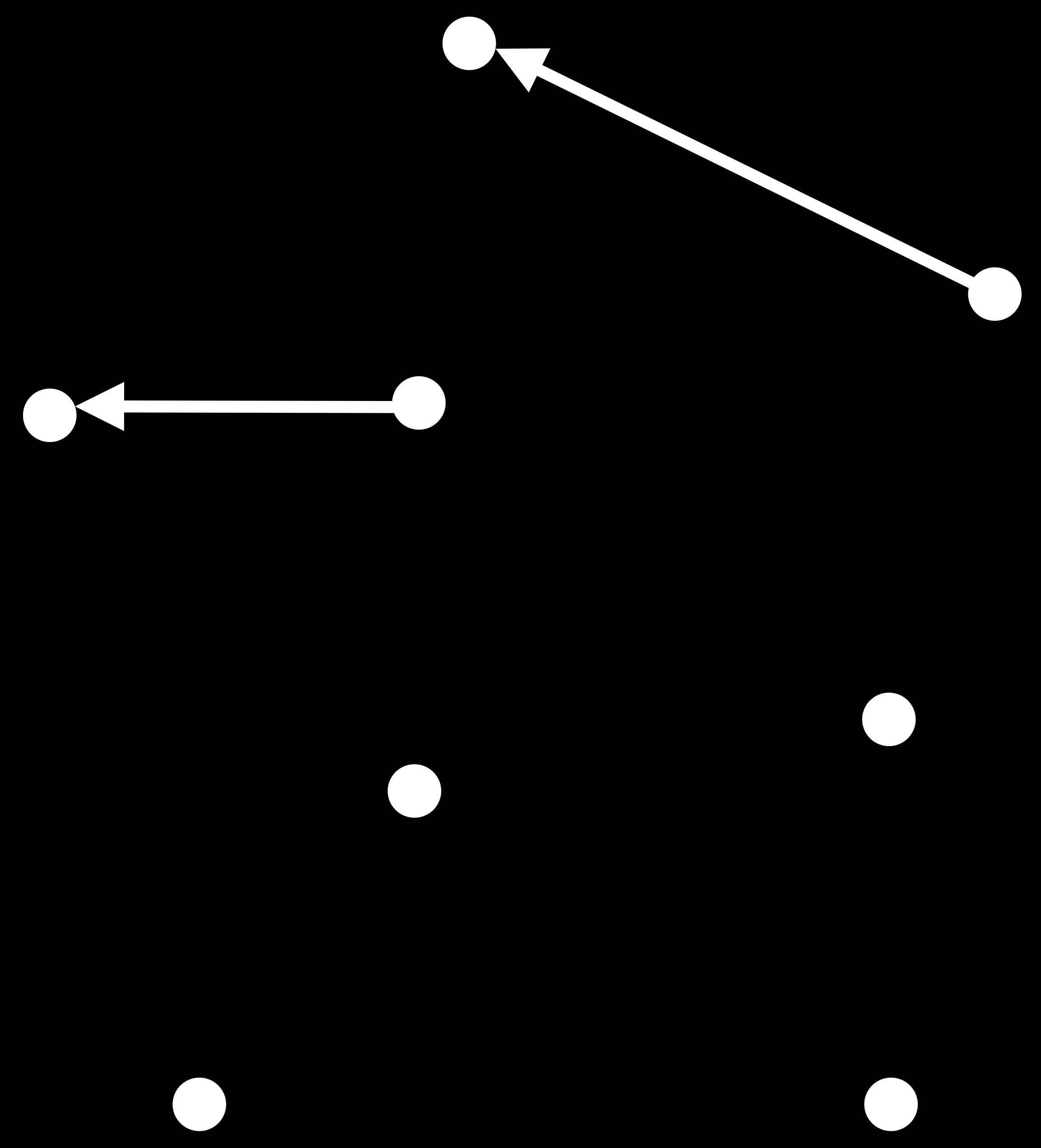
Triadic closure offers small choice sets
→ tractable inference
→ varied choice sets



Our data
Timestamped edges
(including repeats)

Choosing to close triangles

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

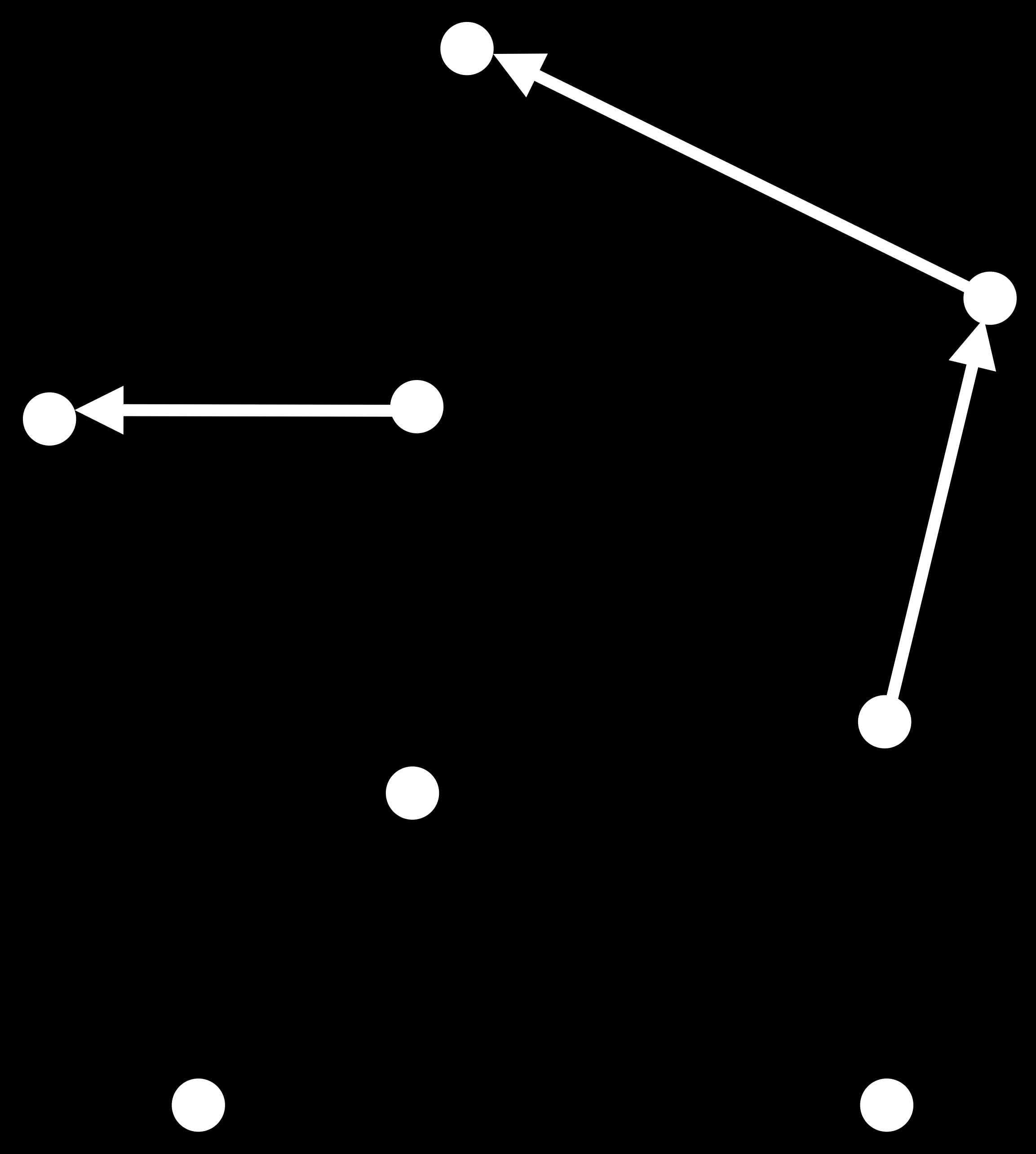


Our data
Timestamped edges
(including repeats)

Choosing to close triangles

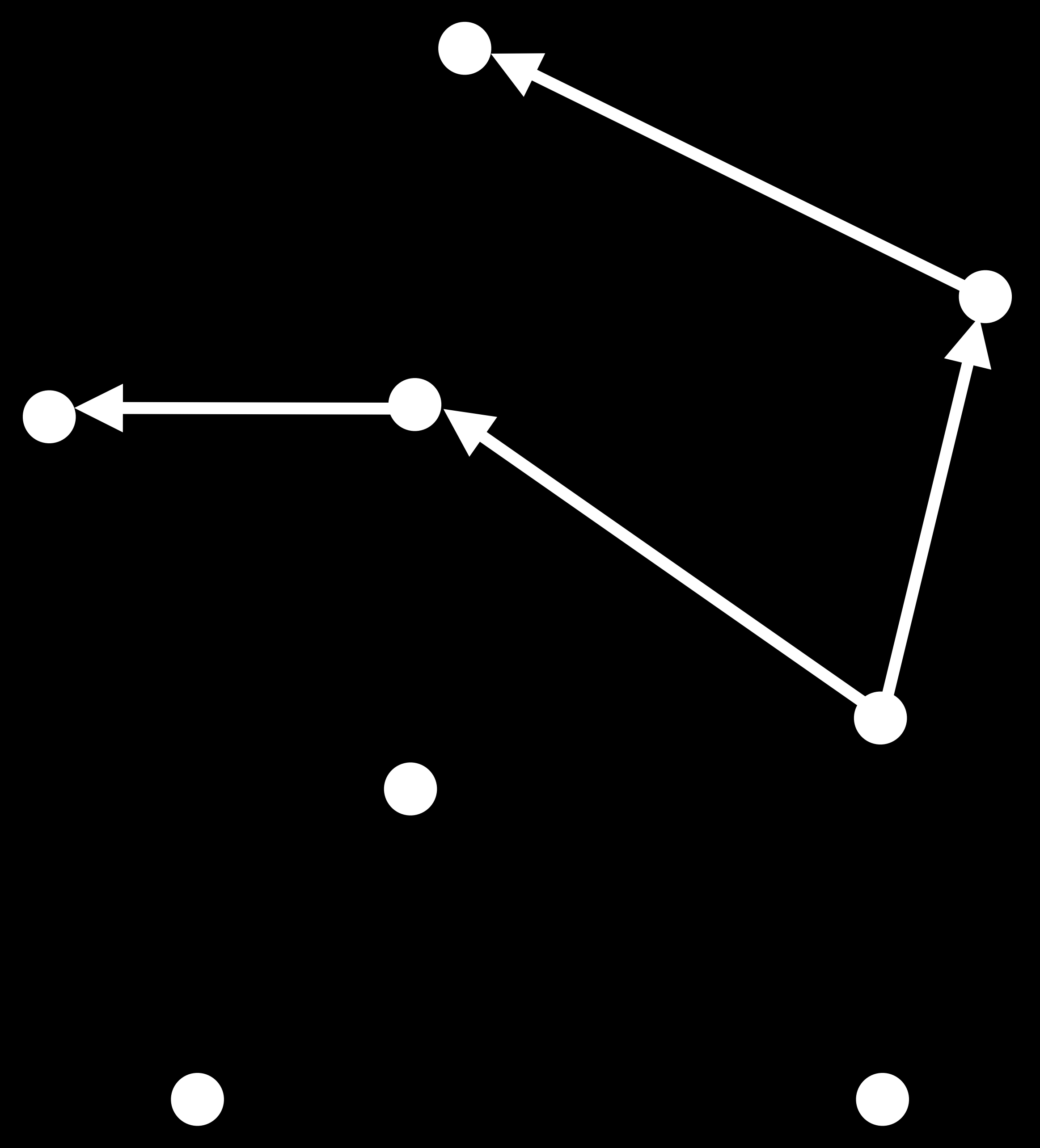
Triadic closure offers small choice sets
→ tractable inference
→ varied choice sets

Our data
Timestamped edges
(including repeats)



Choosing to close triangles

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

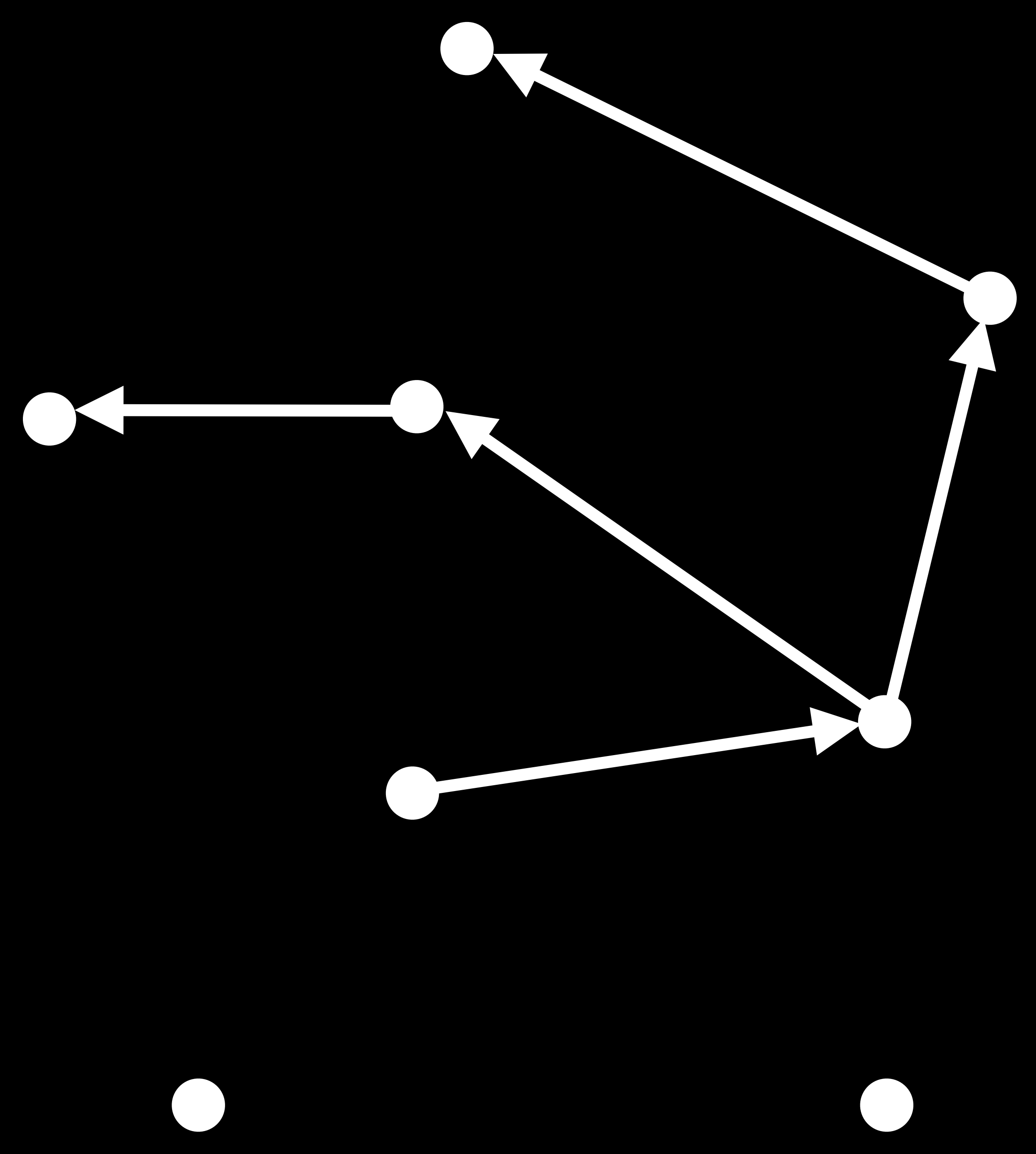


Our data
Timestamped edges
(including repeats)

Choosing to close triangles

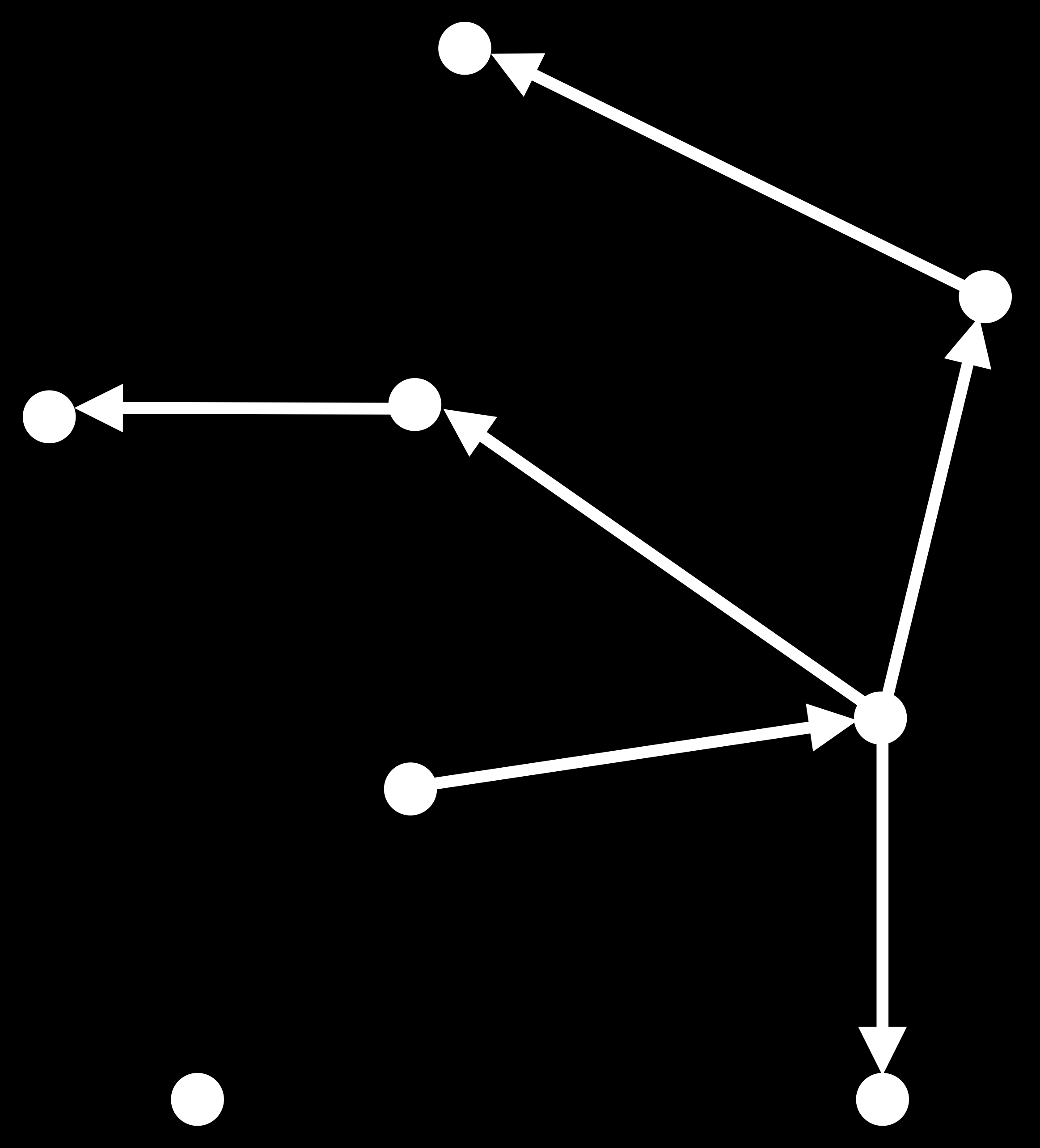
Triadic closure offers small choice sets
→ tractable inference
→ varied choice sets

Our data
Timestamped edges
(including repeats)



Choosing to close triangles

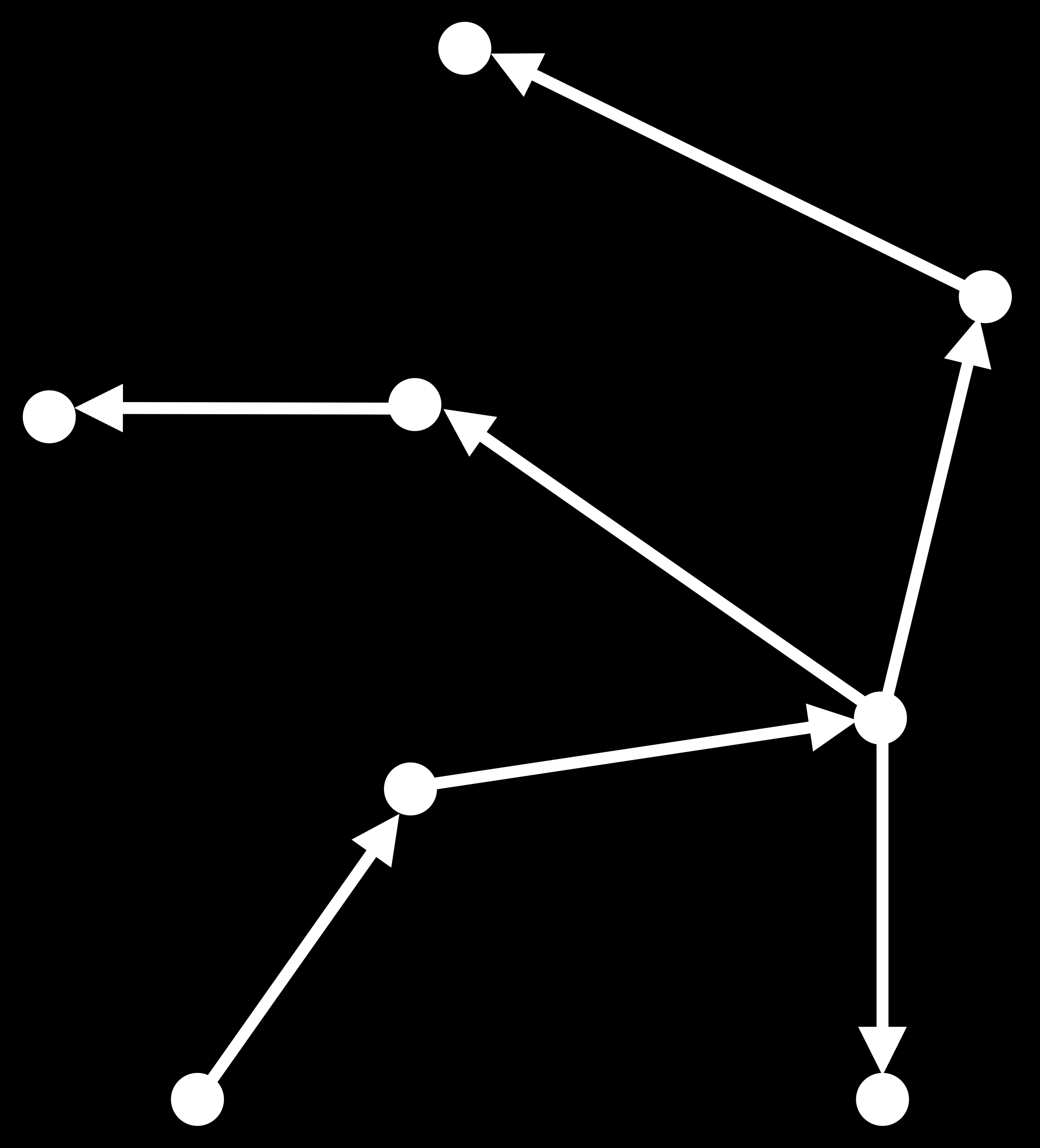
Triadic closure offers small choice sets
→ tractable inference
→ varied choice sets



Our data
Timestamped edges
(including repeats)

Choosing to close triangles

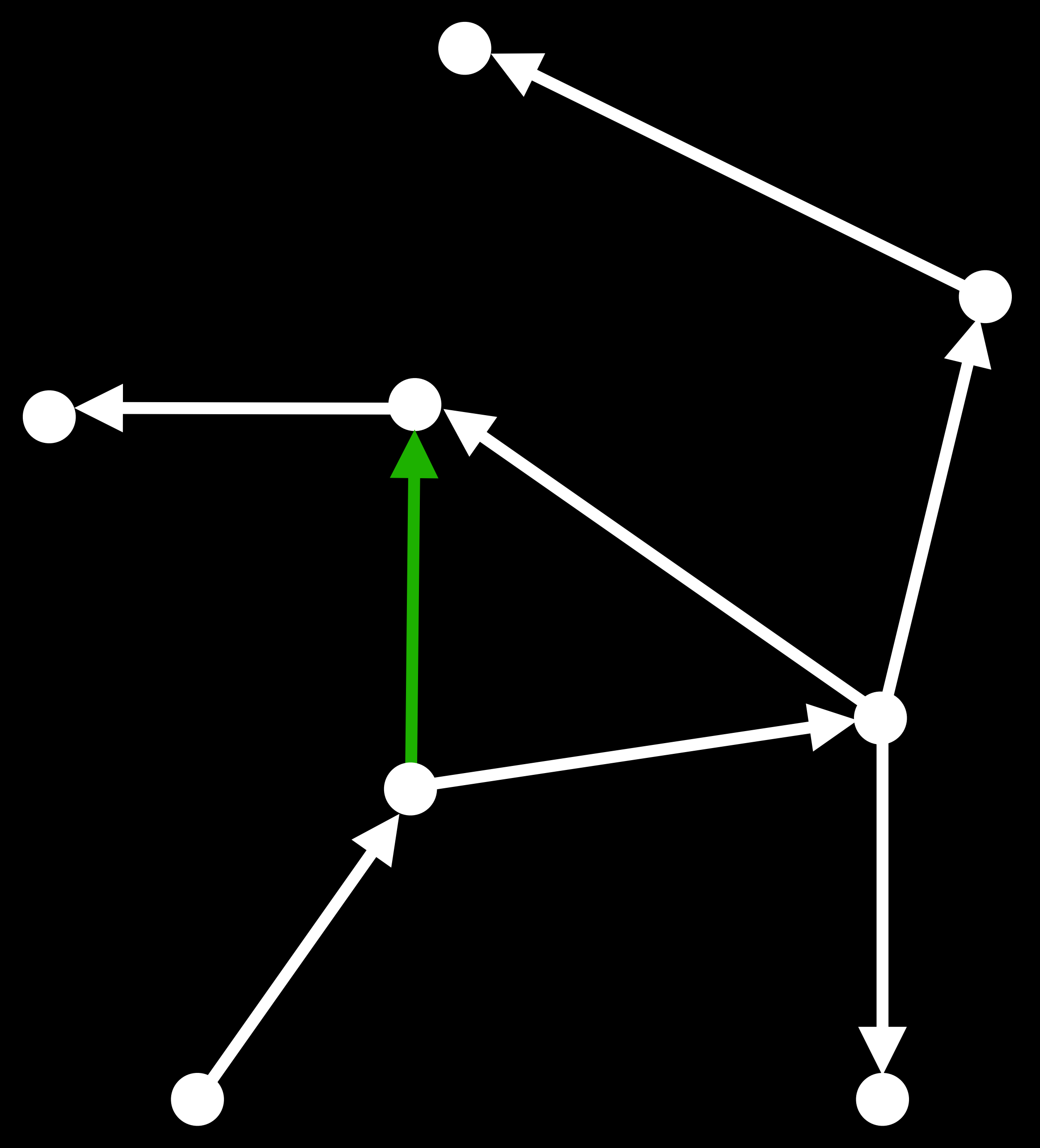
Triadic closure offers small choice sets
→ tractable inference
→ varied choice sets



Our data
Timestamped edges
(including repeats)

Choosing to close triangles

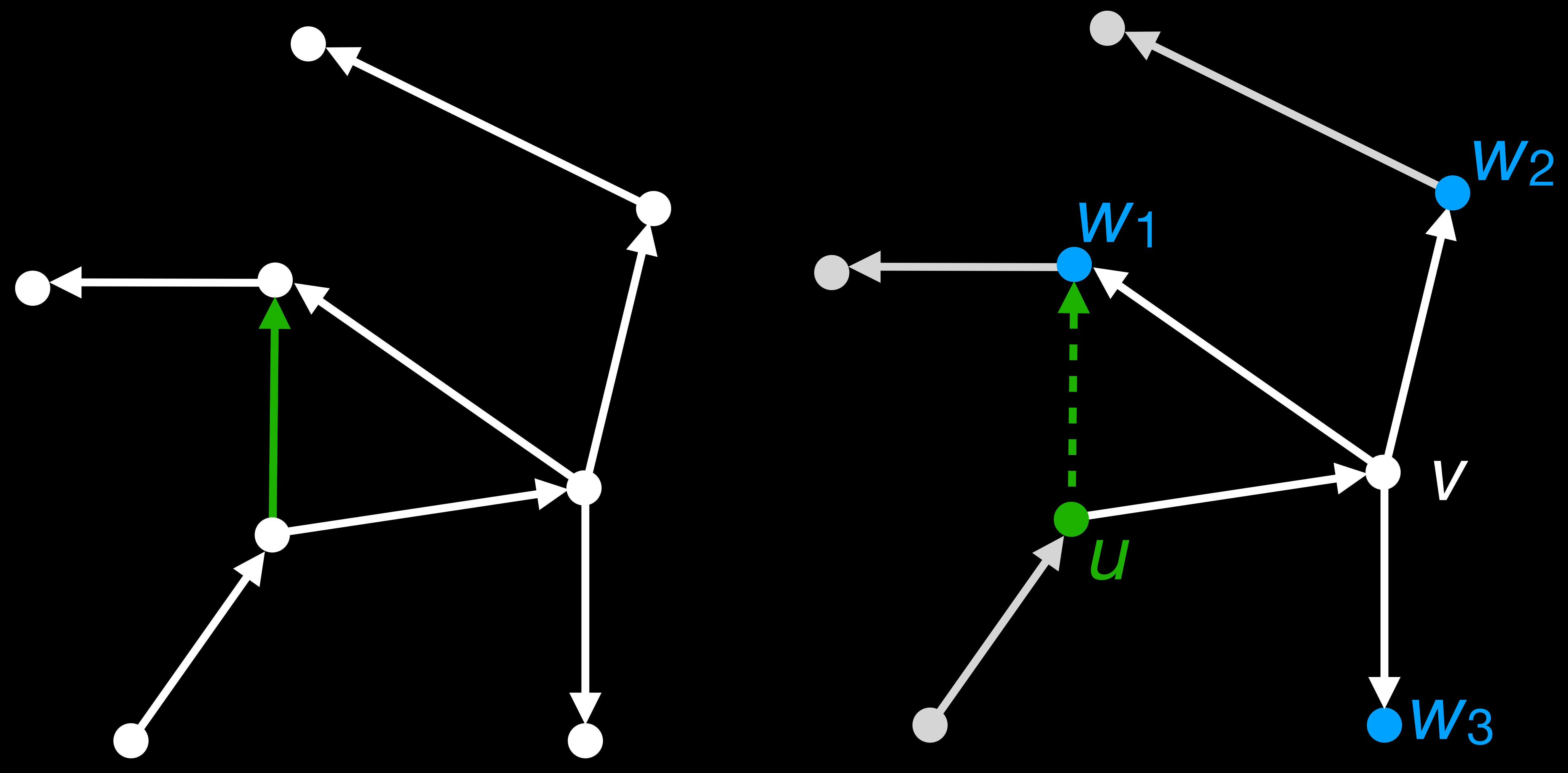
Triadic closure offers small choice sets
→ tractable inference
→ varied choice sets



Our data
Timestamped edges
(including repeats)

Choosing to close triangles

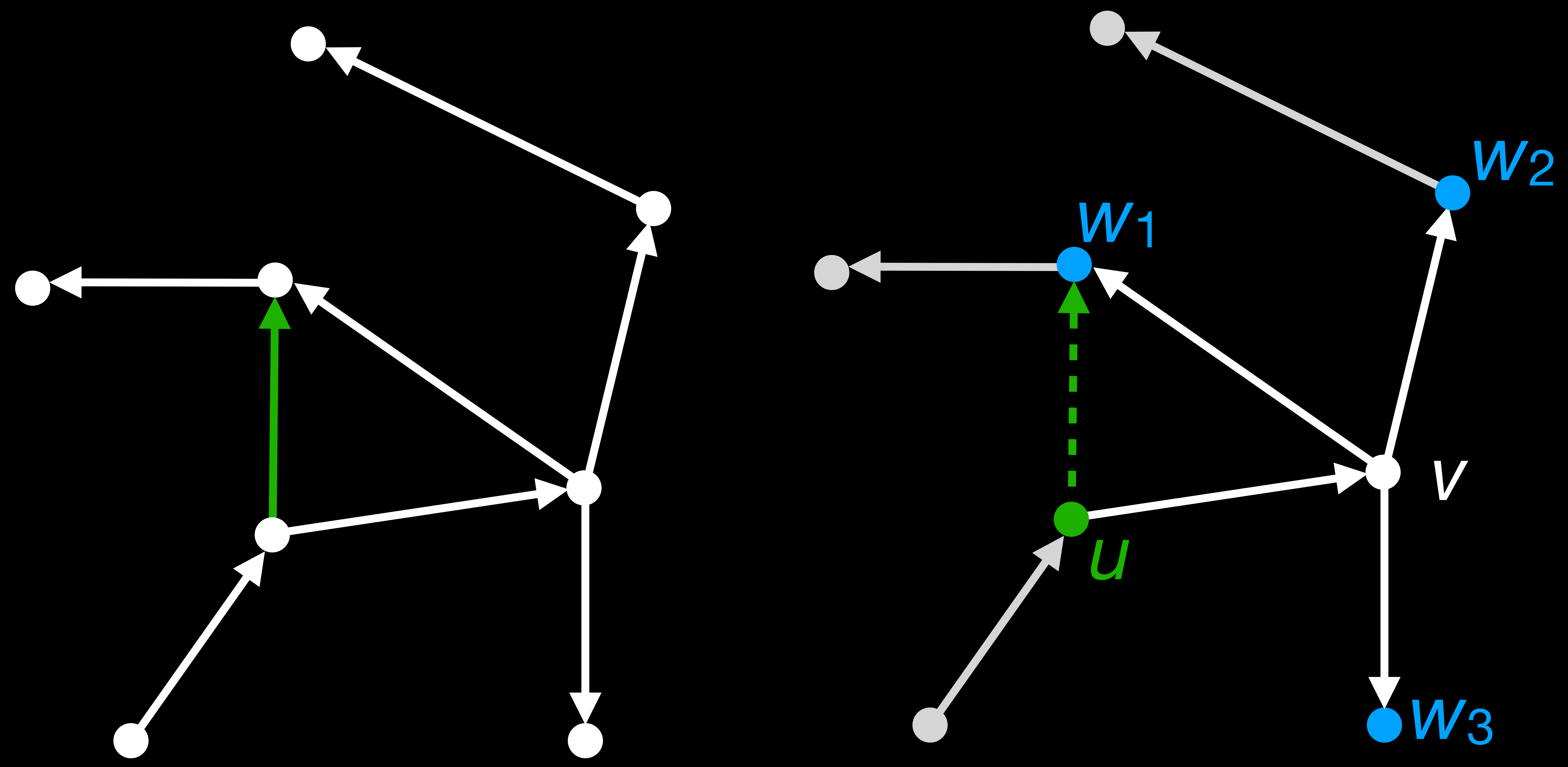
Triadic closure offers small choice sets
→ tractable inference
→ varied choice sets



Our data
Timestamped edges
(including repeats)

Choosing to close triangles

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



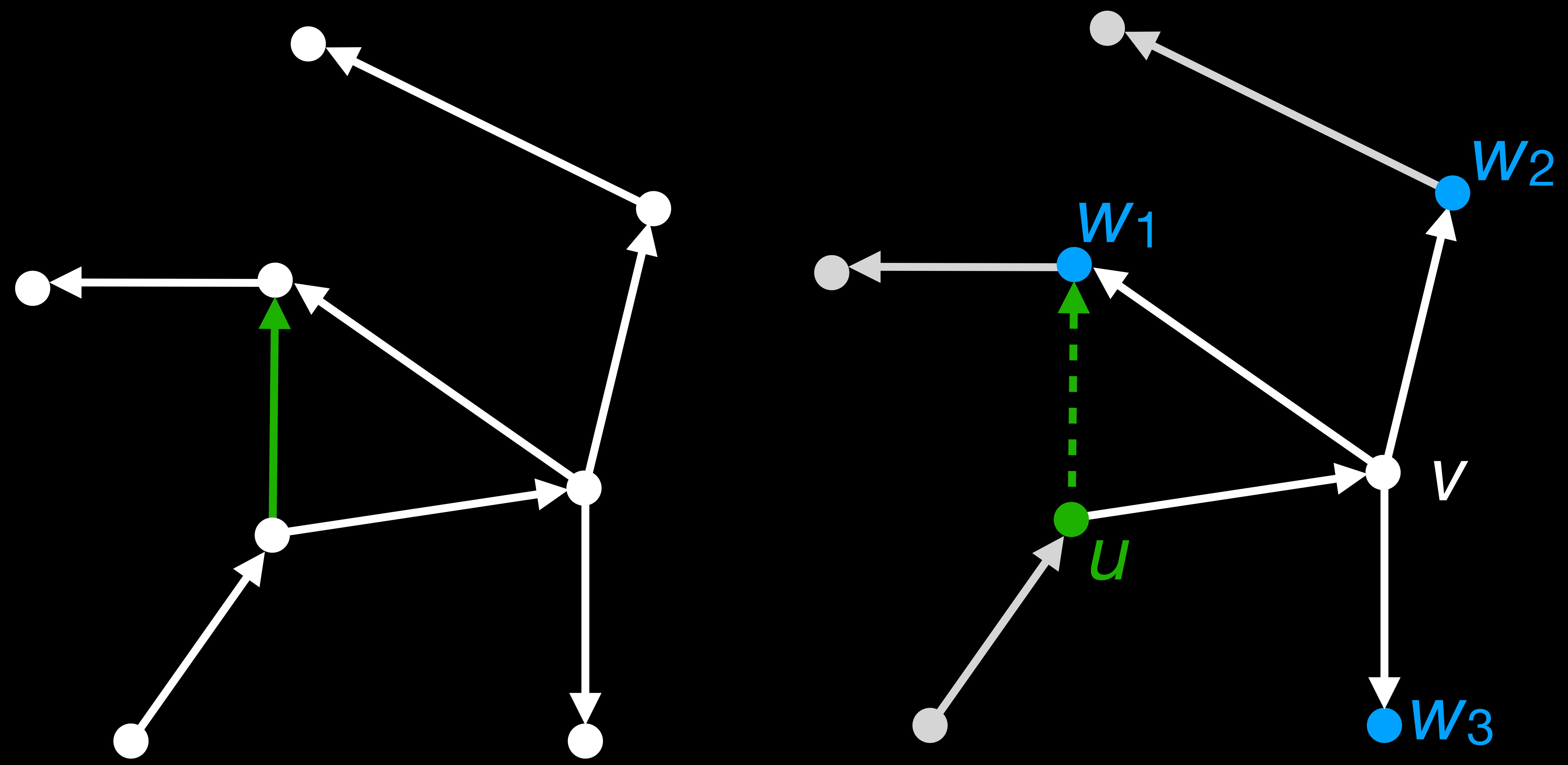
chooser
u

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)



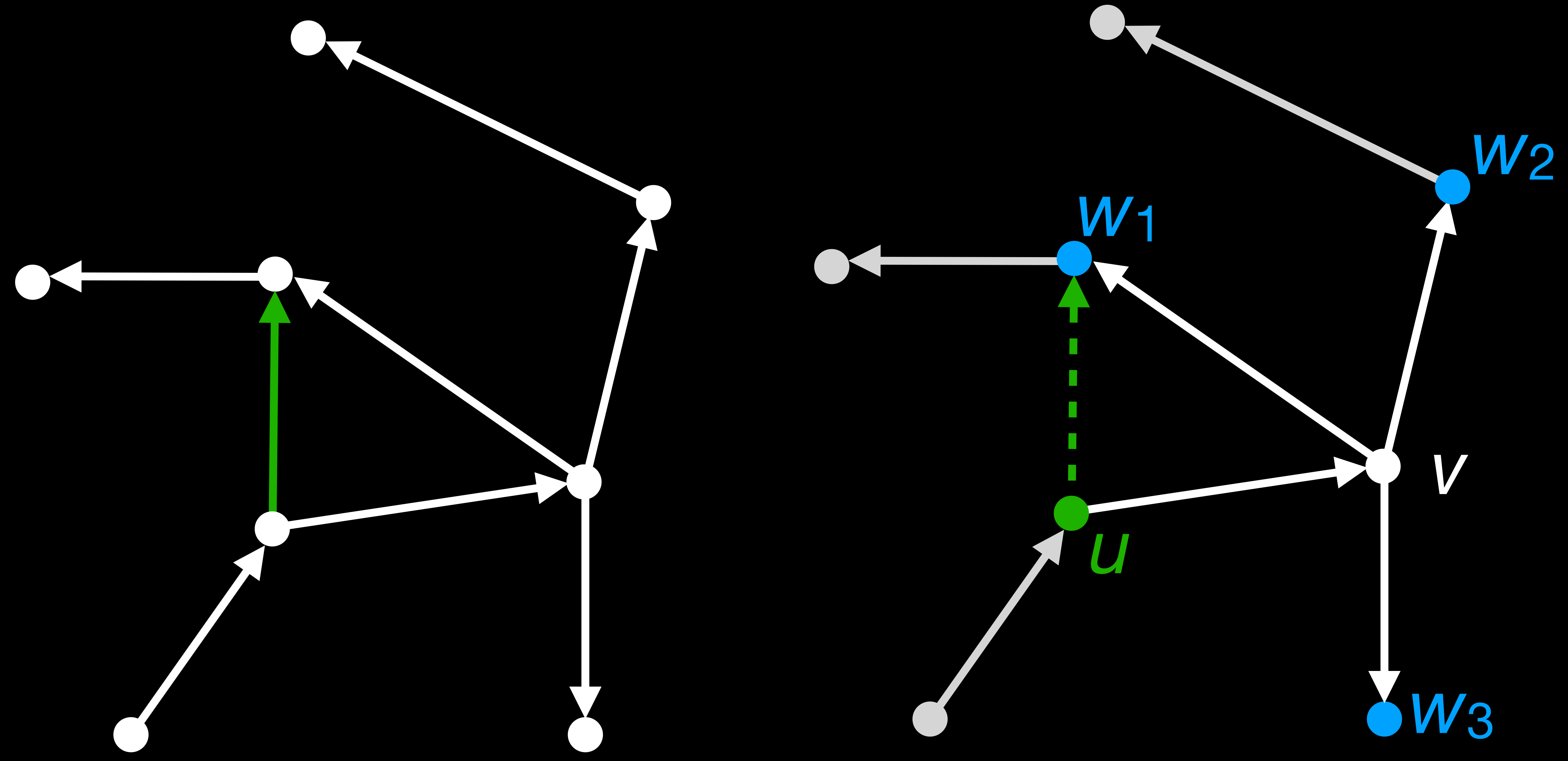
chooser u

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)

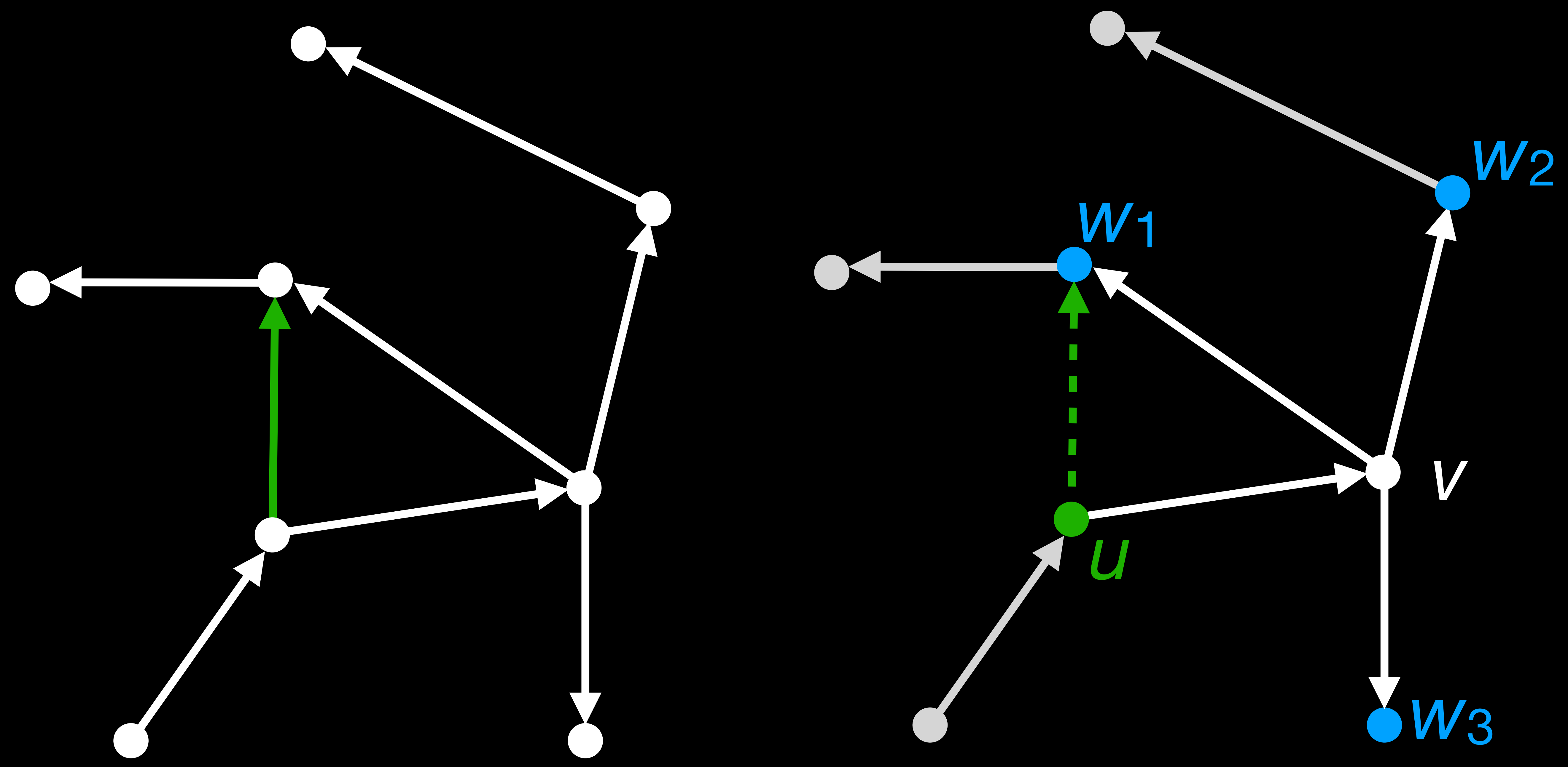


chooser	*choice set*	*choice*
u	$\{w_1, w_2, w_3\}$	w_1

Choosing to close triangles

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

Our data
Timestamped edges
(including repeats)



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

Node features

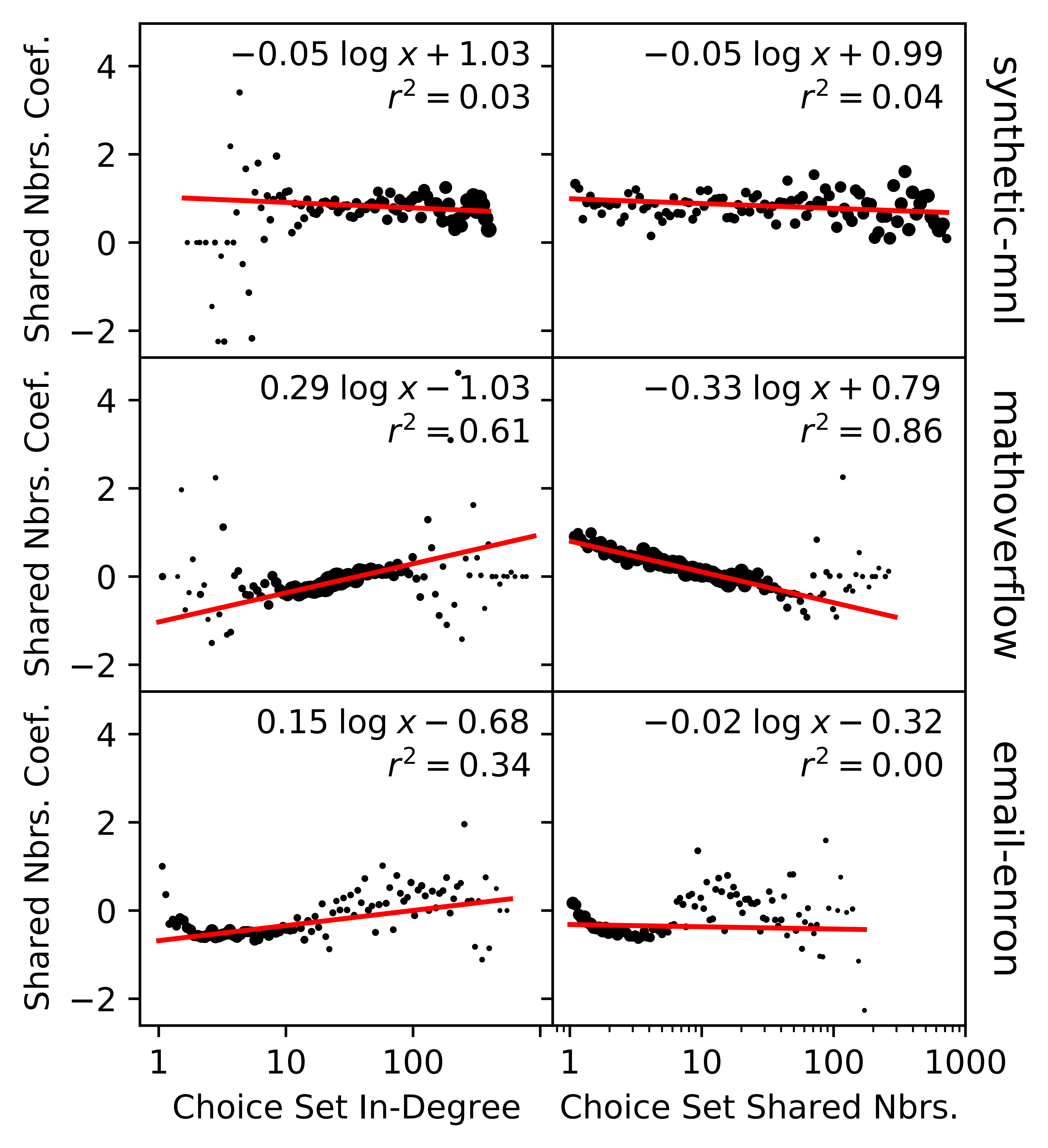
1. in-degree of w
2. # shared neighbors of u, w
3. weight of edge $w \rightarrow u$
4. time since last edge into w
5. time since last edge out of w
6. time since last $w \rightarrow u$ edge

Context matters in triadic closure

Context matters in triadic closure

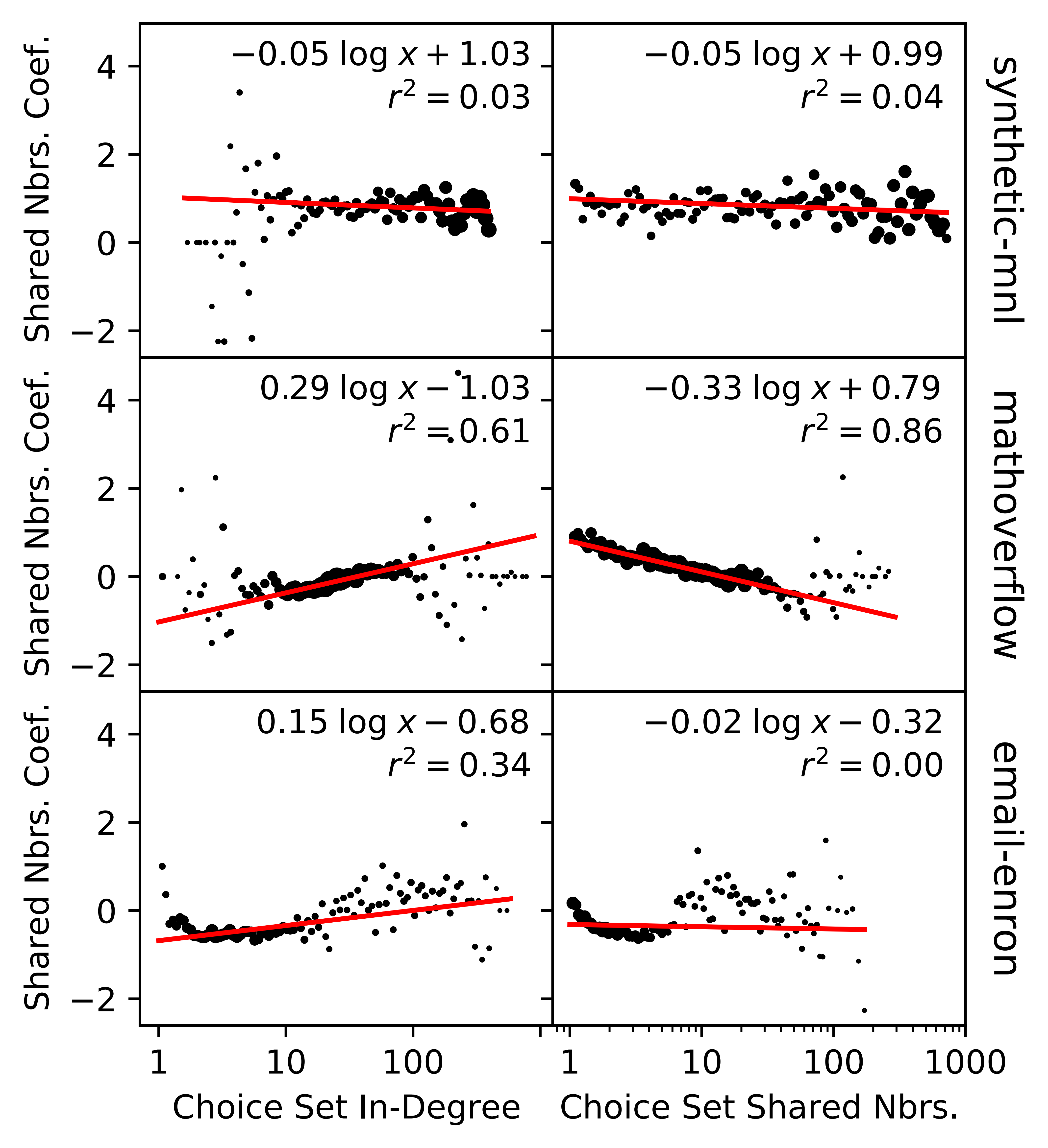
Datasets
email-enron
email-eu
email-w3c
wiki-talk
reddit-hyperlink
bitcoin-alpha
bitcoin-otc
mathoverflow
college-msg
facebook-wall
sms-a
sms-b
sms-c

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

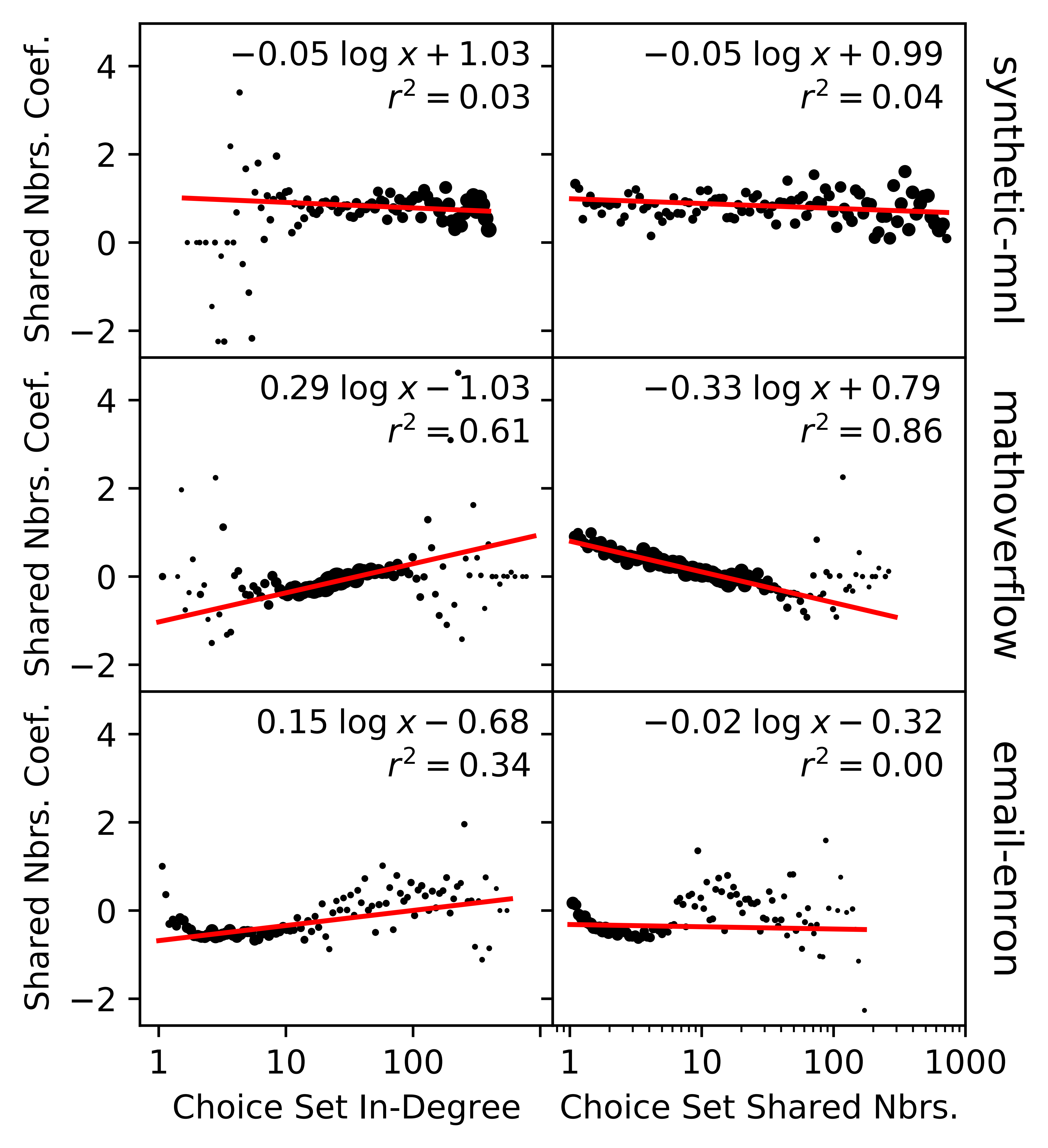
Context matters in triadic closure



Synthetic data,
no context effects

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

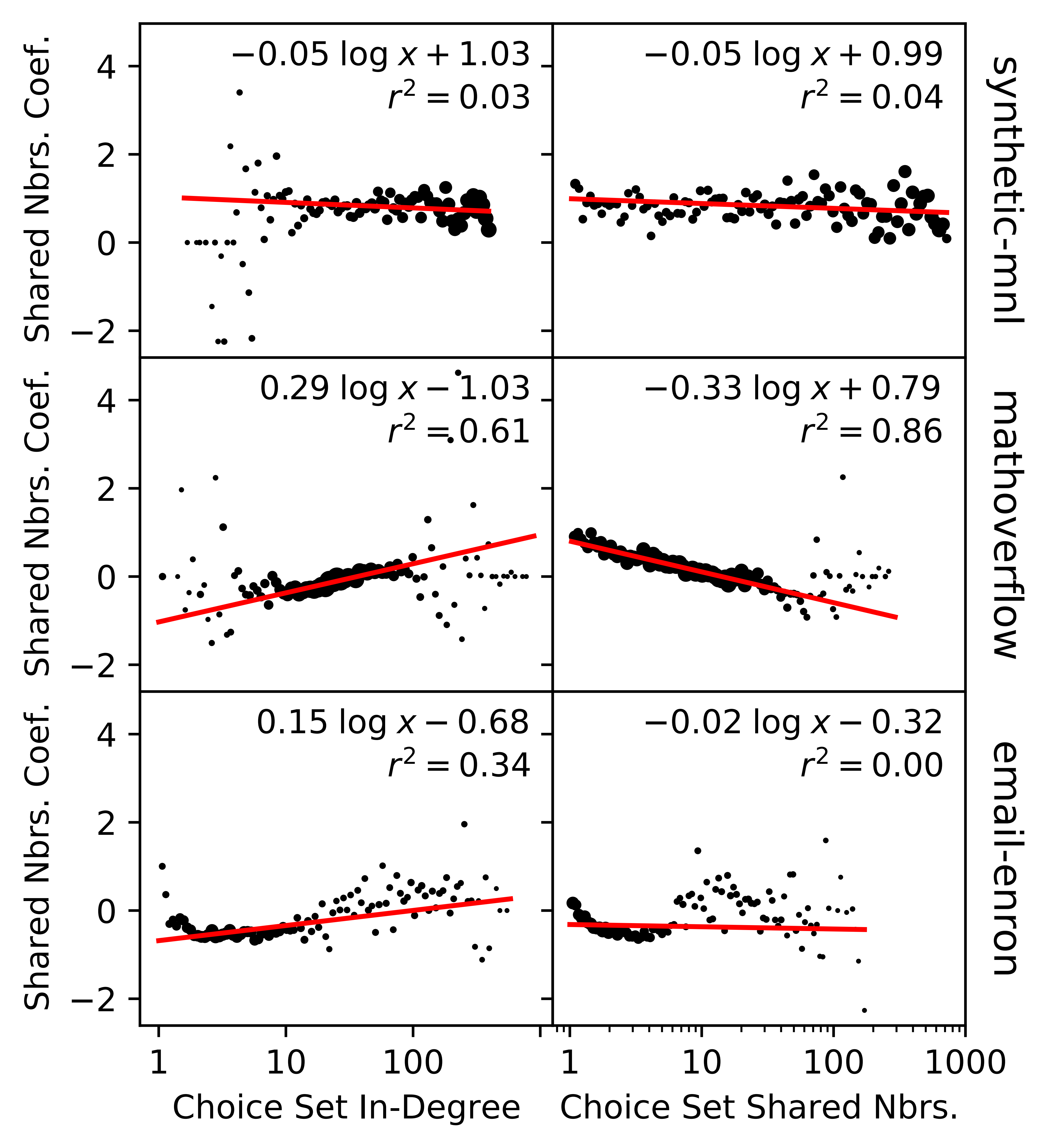


Synthetic data,
no context effects

Commenting network,
linear context effects

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



Synthetic data,
no context effects

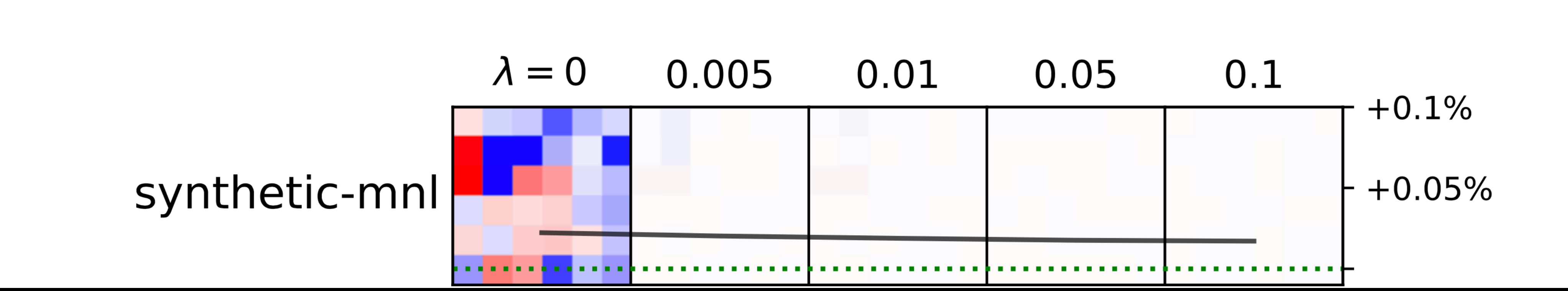
Commenting network,
linear context effects

Email network,
nonlinear context effects?

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

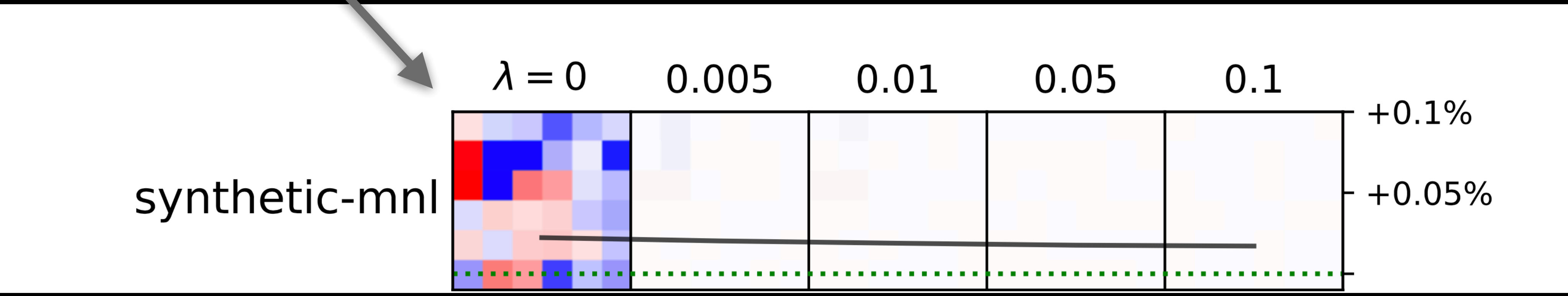
LCL reveals interpretable feature context effects

LCL reveals interpretable feature context effects



LCL reveals interpretable feature context effects

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



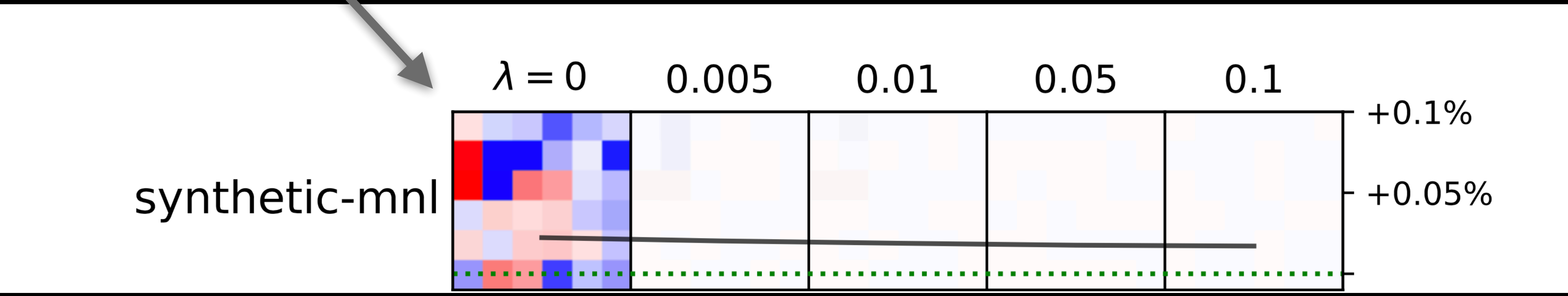
Node features
(left-right, top-bottom)

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

LCL reveals interpretable feature context effects

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

increasing L_1 regularization on A



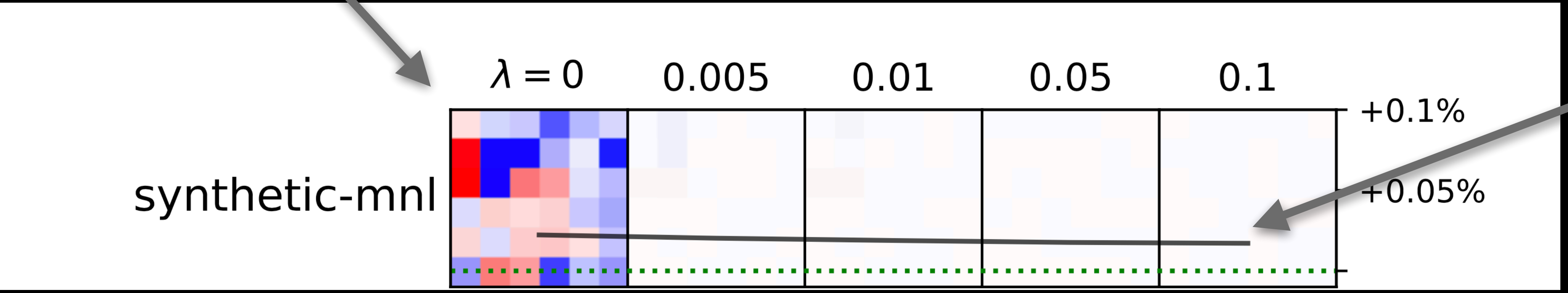
Node features
(left-right, top-bottom)

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

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)

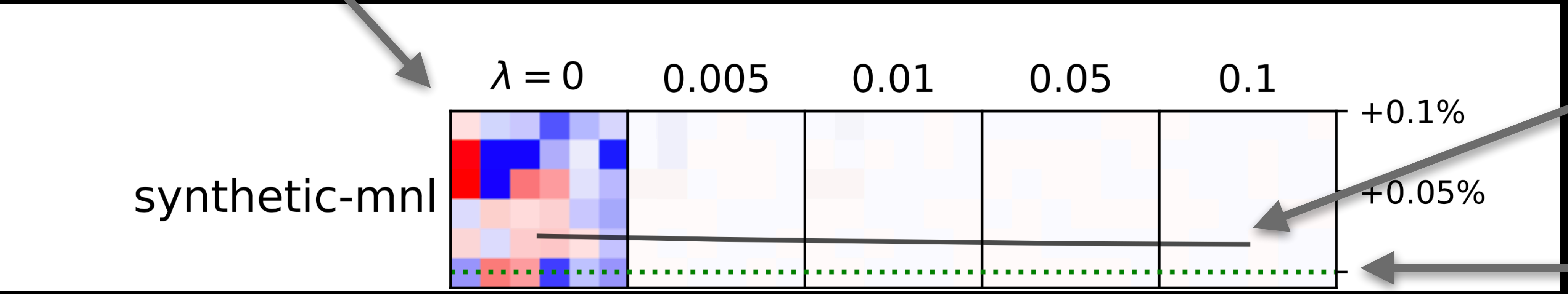
Node features
(left-right, top-bottom)

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

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

Node features
(left-right, top-bottom)

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

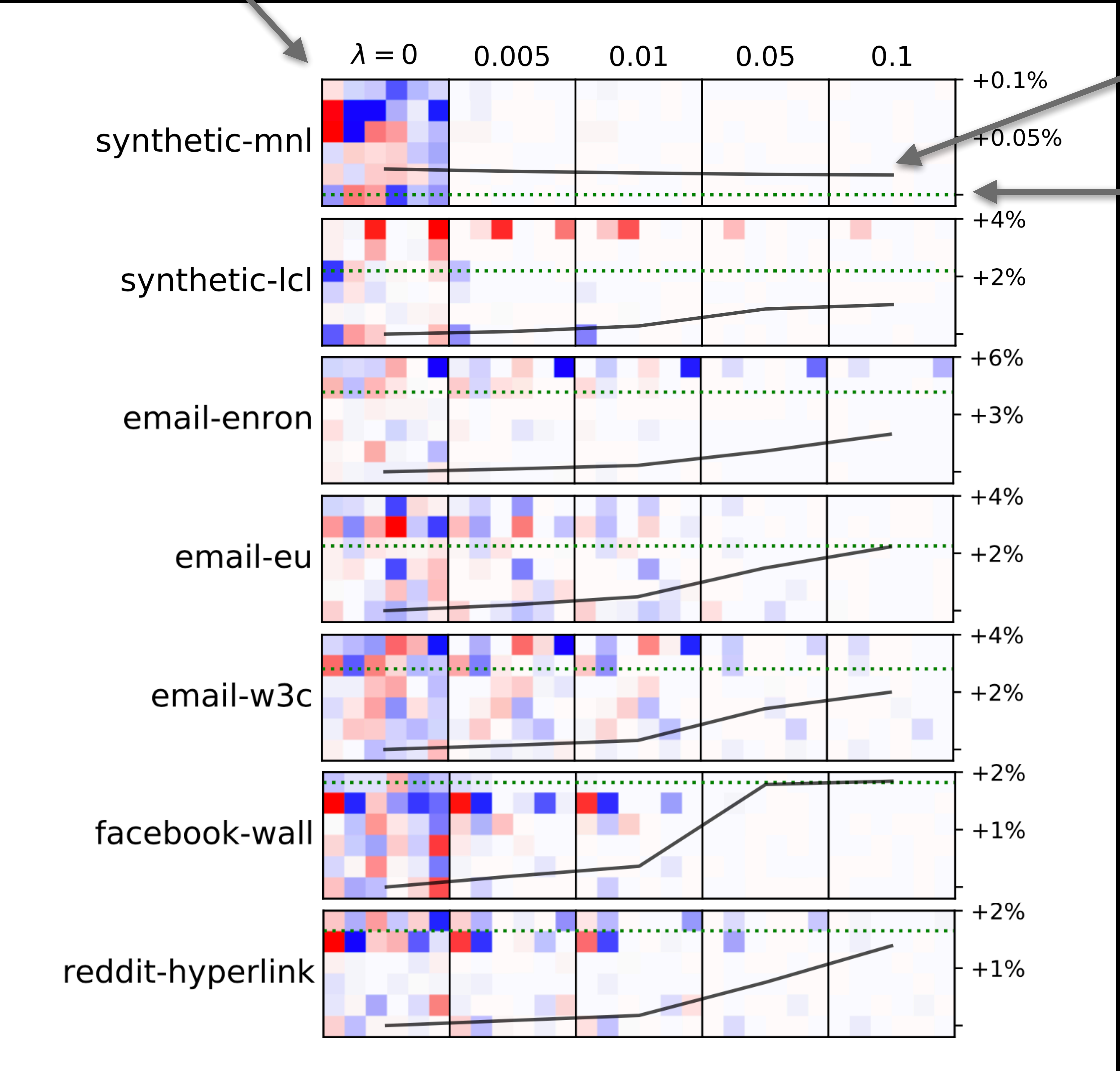
LCL reveals interpretable feature context effects

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



Node features
(left-right, top-bottom)

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

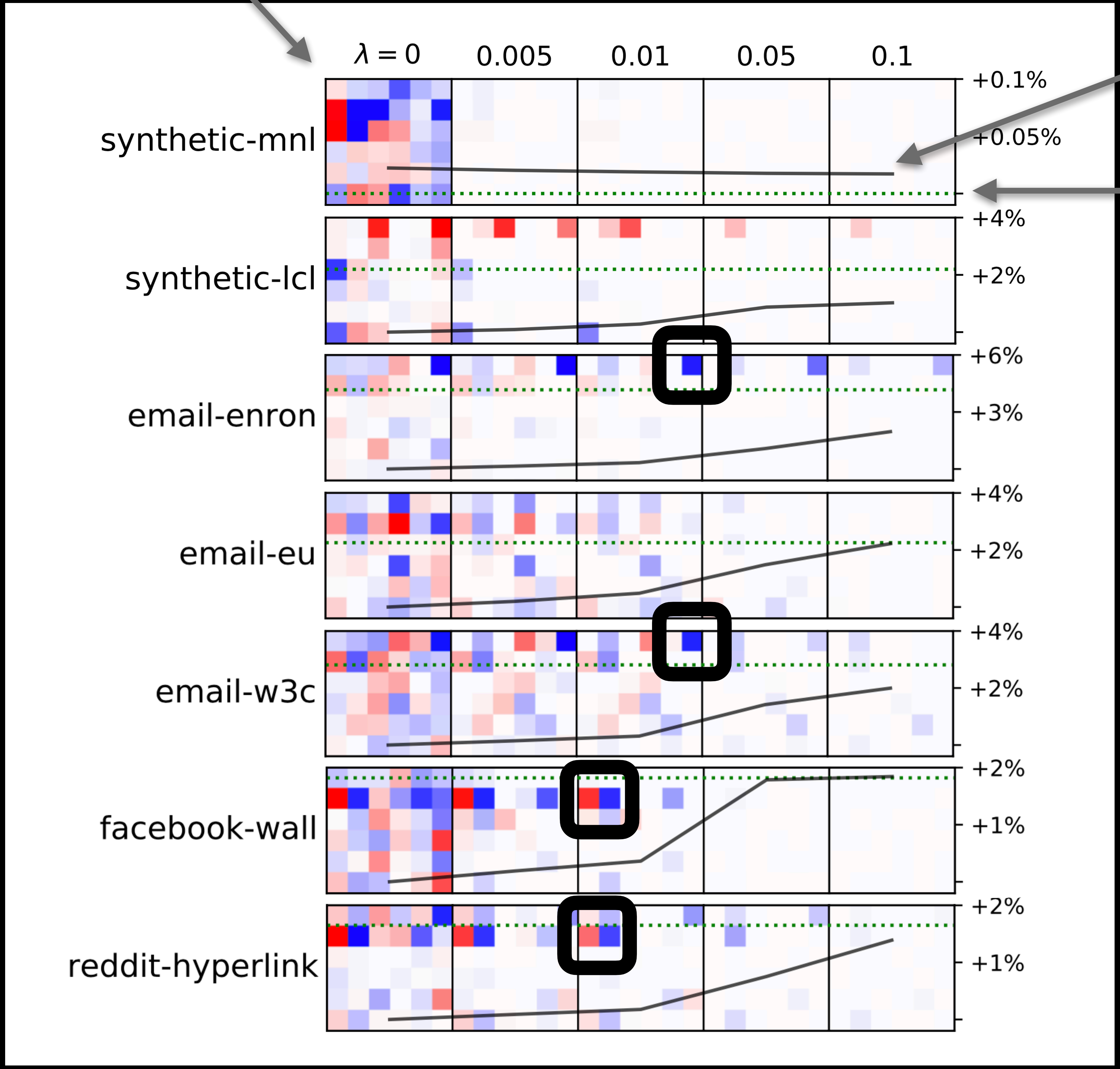
LCL reveals interpretable feature context effects

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



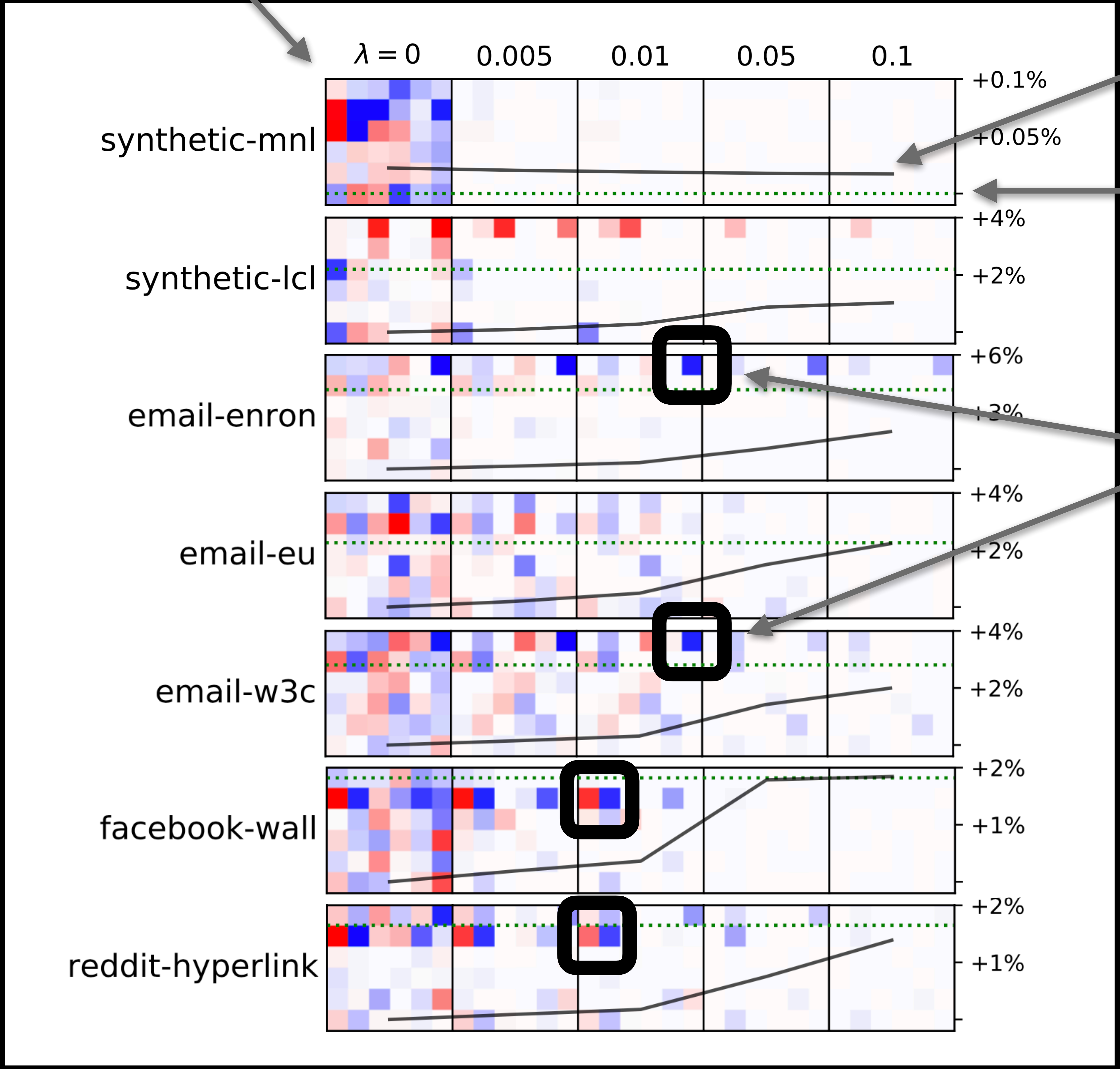
Node features
(left-right, top-bottom)

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

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

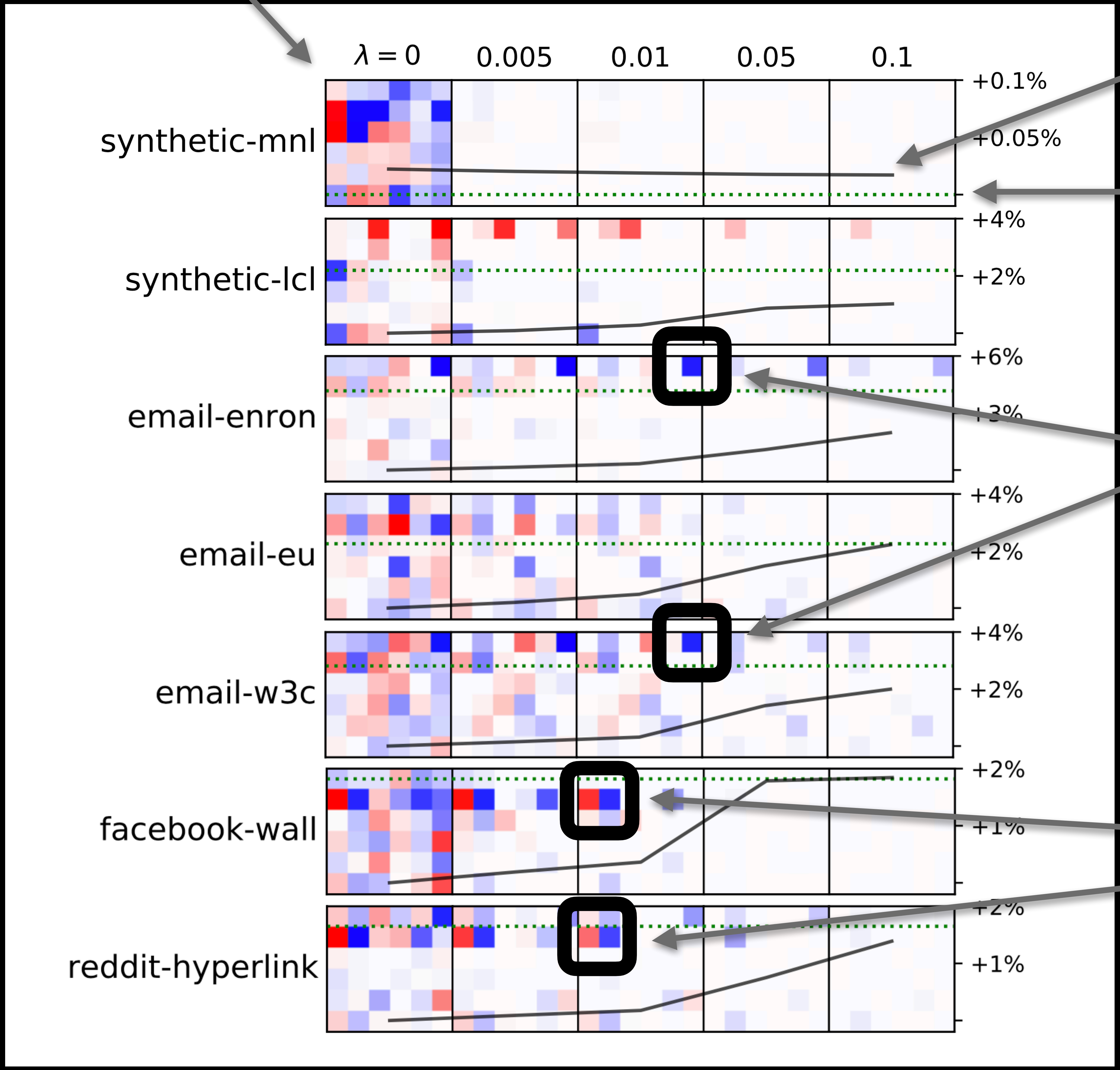
Node features
(left-right, top-bottom)

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

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

Node features
(left-right, top-bottom)

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

"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

Code: bit.ly/lcl-code
Data: bit.ly/lcl-data
Slides: bit.ly/lcl-kdd-slides

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”

Submit to our NeurIPS ’21 workshop!
bit.ly/WHMD2021

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