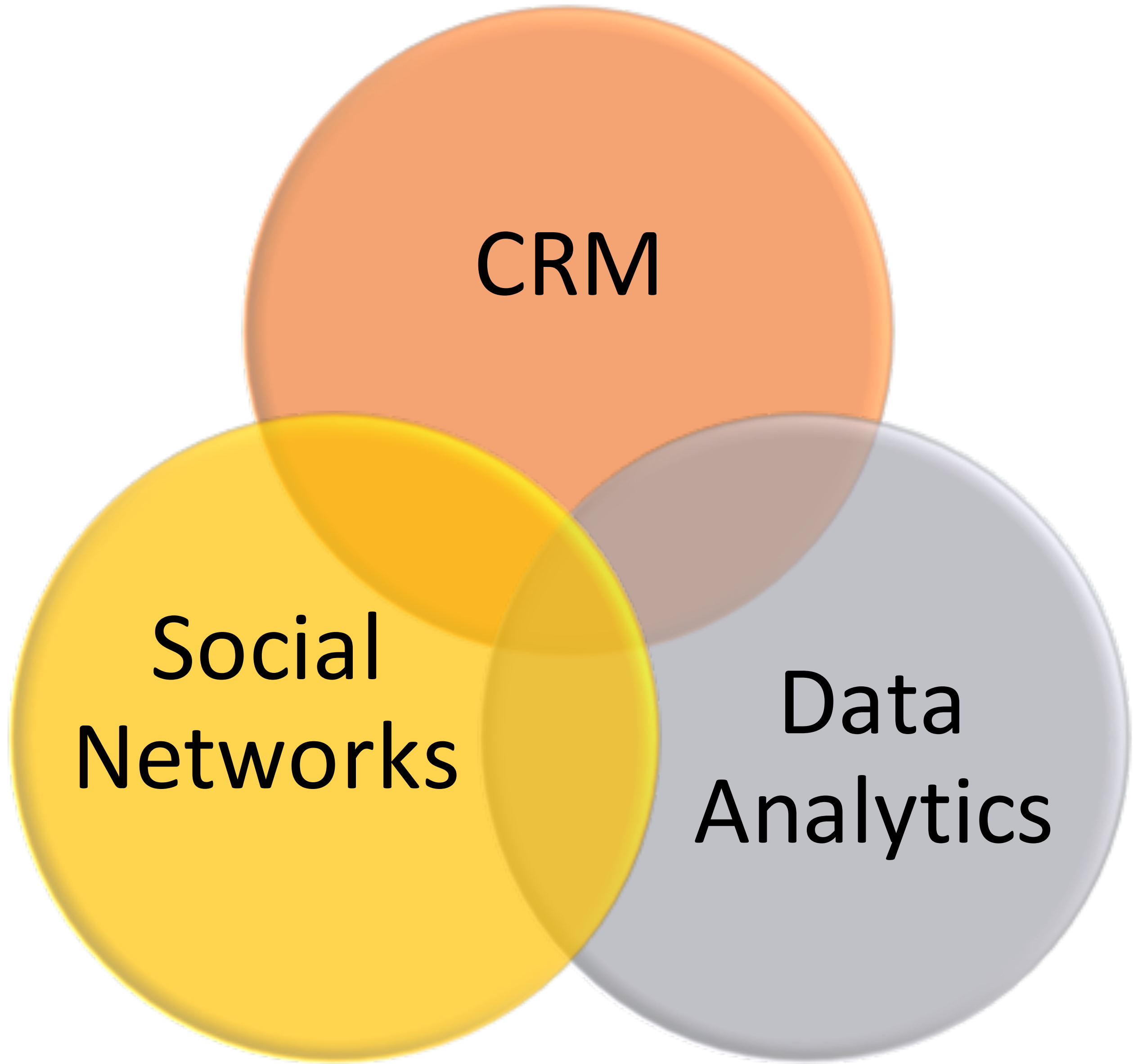
UCD STATISTICS Talk

Pantelis Loupos
Graduate School Of Management
University Of California, Davis

MY RESEARCH INTERESTS



Methods: Predictive Analytics (Machine Learning, Data Mining)

Causal Inference (Matching, Instrumental Variables, A/B Testing)

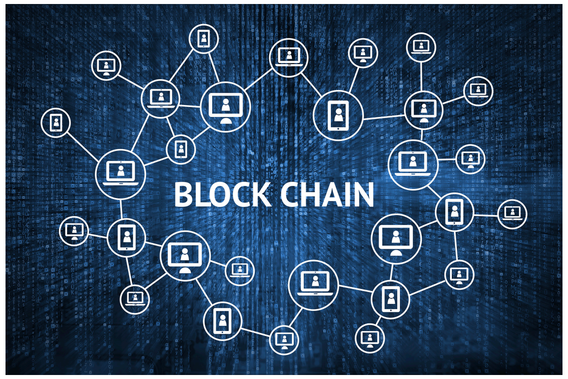
SOCIAL PLATFORMS

- **Social Platforms/Services:** users create value for other users!
- Dimension: **Networks Externality**

UBER

airbnb

venmo

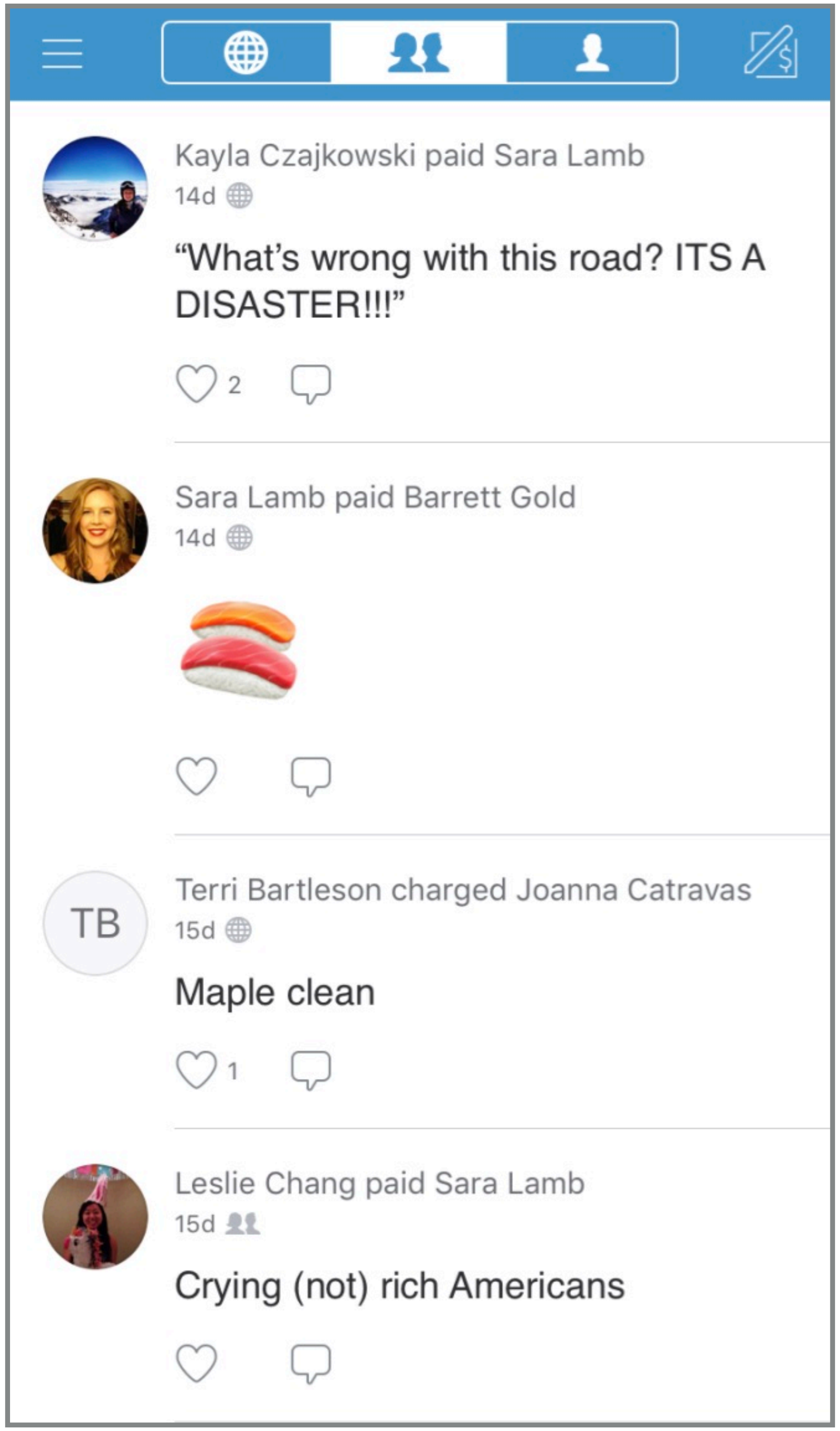


Starting Cold: The Power Of Social Networks In Predicting Non-Contractual Customer Behavior

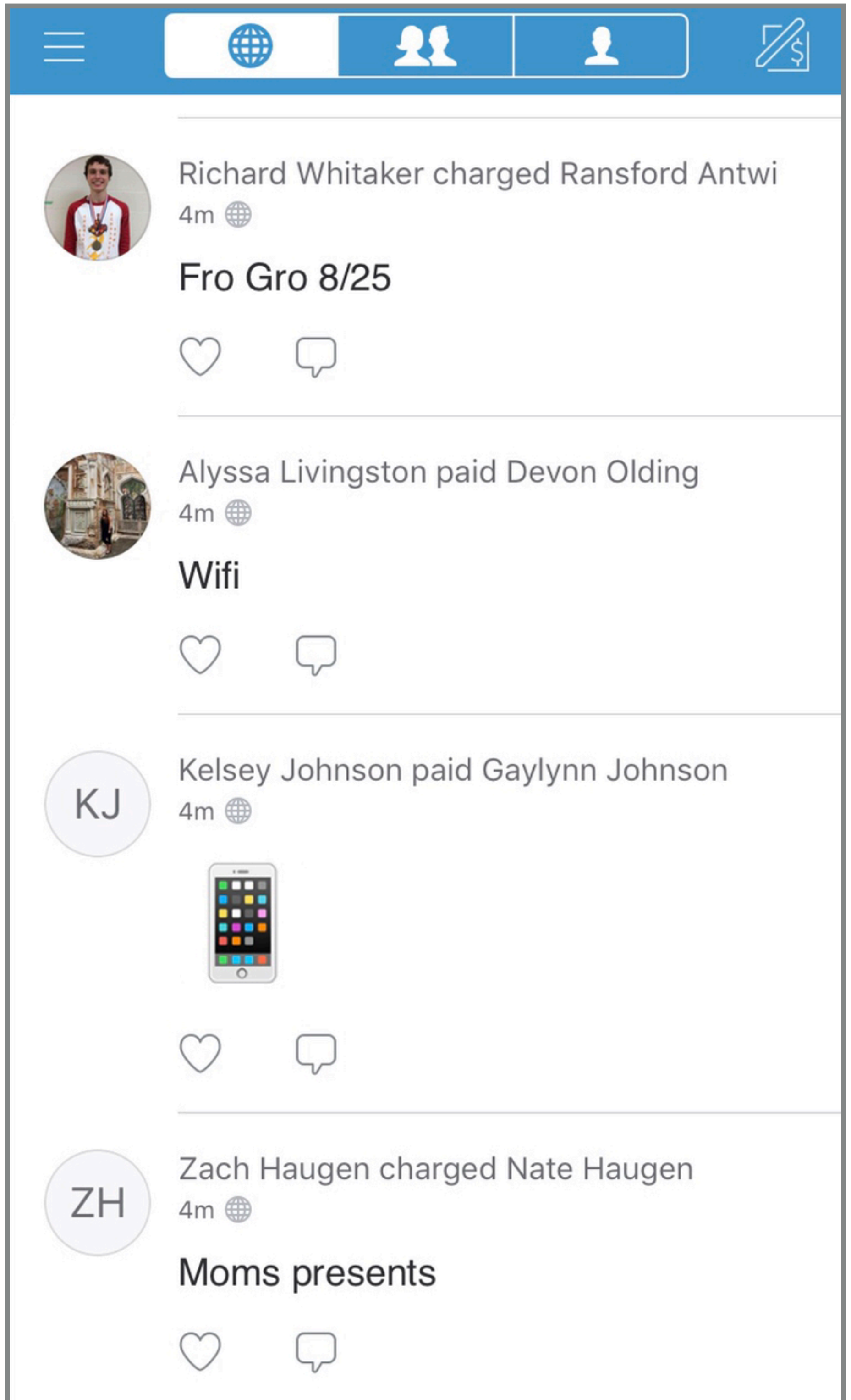
Joint Work With Alexandros Nathan And Moran Cerf

ALICE DISCOVERS VENMO

VENMO IS SOCIAL



VENMO IS SOCIAL



- Transformed payments into a sharing experience!
- Largest P2P Financial Transaction Network
  - 16+M Users
  - 10M **Active** Monthly users
  - Q4 2018: $**19B**
- Fast Growth Company

PROBLEM MOTIVATION: WHO IS WORTH IT?

- Alice just joined Venmo.
- Should Venmo invest its marketing dollars:
  - **with or without behavioral data?**
- This is the **cold-start** **problem**.

PROBLEM MOTIVATION: WHO IS WORTH IT?

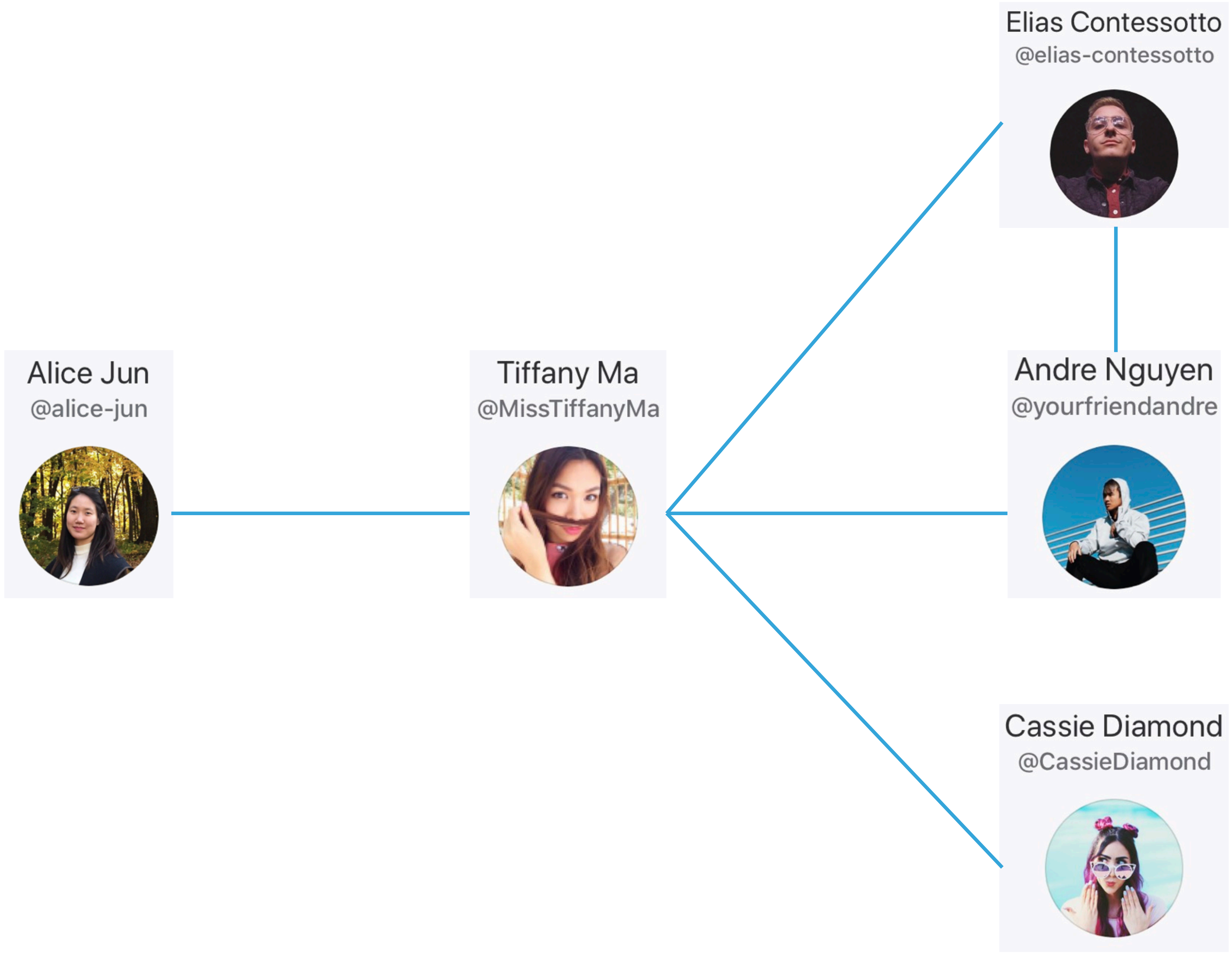
Can social network data help solve this?

- **with or without behavioral data?**

- This is the **cold-start problem**.

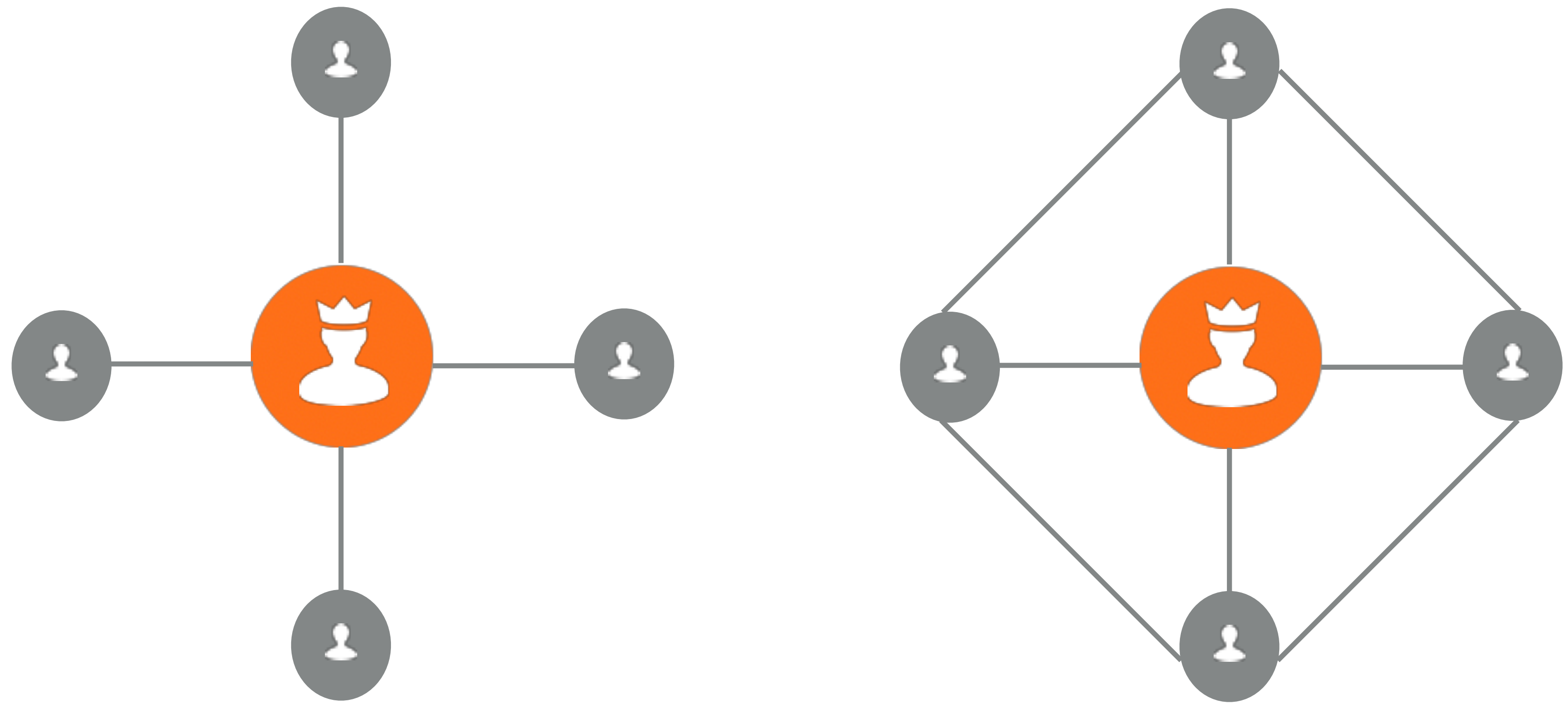
INTUITION

"Tell Me Your Friends And I Will Tell You What Type Of Customer You Are "



WHY SHOULD ALICE'S FRIENDS ALLOW VENMO TO INFER HOW GOOD A CUSTOMER SHE IS?

1. Homophily (e.g., Aral et al. 2009)
2. Peer Influence (e.g., Iribarren and Moro 2009)
3. Bott's theory (Bott 2014)



- The purpose of this work is purely **predictive**

NON-CONTRACTUAL PREDICTIONS ARE CHALLENGING

- Who has churned?
  - No formal declaration of termination (Ascarza et al. 2017)

THIS PAPER CONTRIBUTES TO THE LITERATURE OF COLD START, SOCIAL NETWORKS AND CRM

- Cold Start
  - Recommender Systems (Jamali and Ester 2010)
  - Research Output (Ductor et al. 2014)
- Social Connectivity
  - Neighbor churn (Dasgupta et al. 2008)
  - Social network connections/embeddedness (Benedek et al. 2014)
  - Neighbor/connections usage (Ascarza et al. 2017)
- Non-Contractual CRM
  - Pareto/NDB model (Morrison and Schmittlein 1988; Jerath et al. 2011)
  - BG/NBD extension (Fader et al. 2005; Fader and Hardie 2009)

THIS PAPER CONTRIBUTES TO THE LITERATURE OF CRM, SOCIAL NETWORKS AND COLD START

- Cold Start & Non-Contractual CRM

1. **Acquisition** Related Variables

   - Acquisition Channel (Verhoef and Donkers 2005; Villanueva, Yoo, and Hanssens 2018; Chan, Wu, and Xie 2011; Steffes, Murthi, and Rao 2011)

   - Acquisition Strategy (Steffes, Murthi, and Rao 2011; Lewis 2006; Schmitt, Skiera, and Van den Bulte 2011; Uncles, East, and Lomax 2013; Datta, Foubert, and Van Heerde 2015)

2. **Transactional** Variables

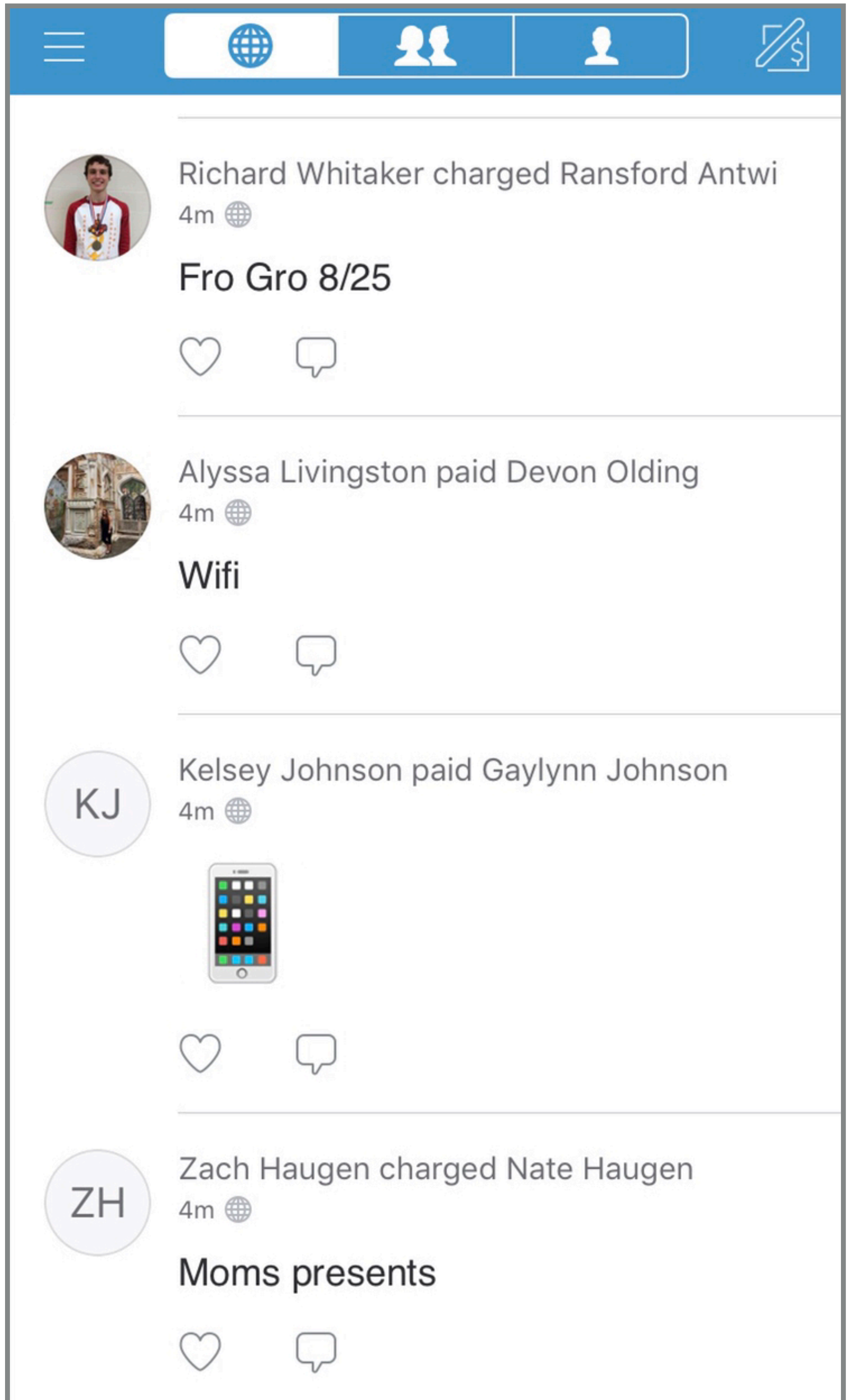
   - Cross-Cohort Changepoint Model (Gopalakrishnan et al. 2017): BYTD model – “borrows” information from previous cohorts

   - First Impressions Count (Padilla and Ascarza 2017): Use the features of a customer’s first transaction

RESEARCH QUESTIONS

- Can we solve the **cold-start** problem?
- Power of Social Networks in Predicting:
    1. Customer Activity
    2. Transaction Volume
    3. 10% Most Frequent Customers

VENMO DATA



- Crawl Venmo's API:
  - **2.3M** Public Users
  - **120M** Financial Transactions

VENMO METRICS

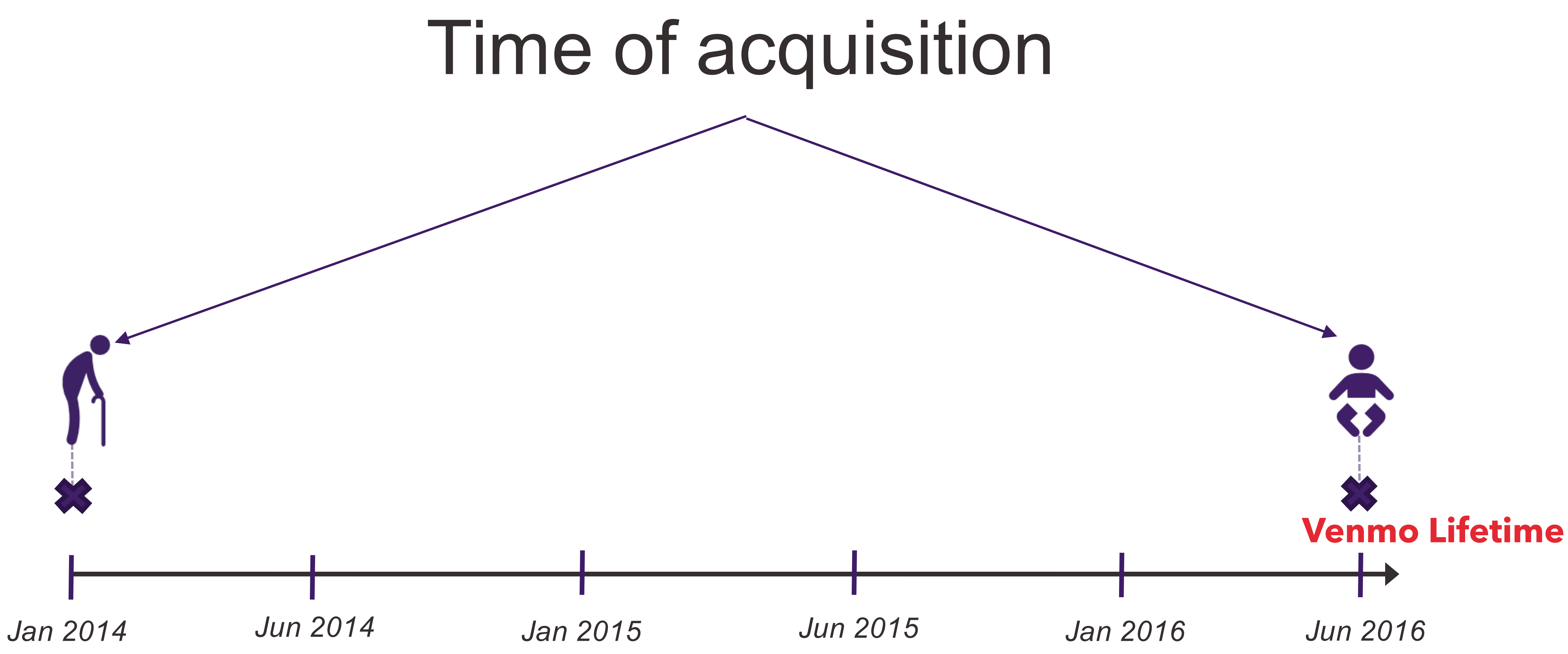
User Based

- Recency
- Frequency
- FB Sign-Up

Social Network Based

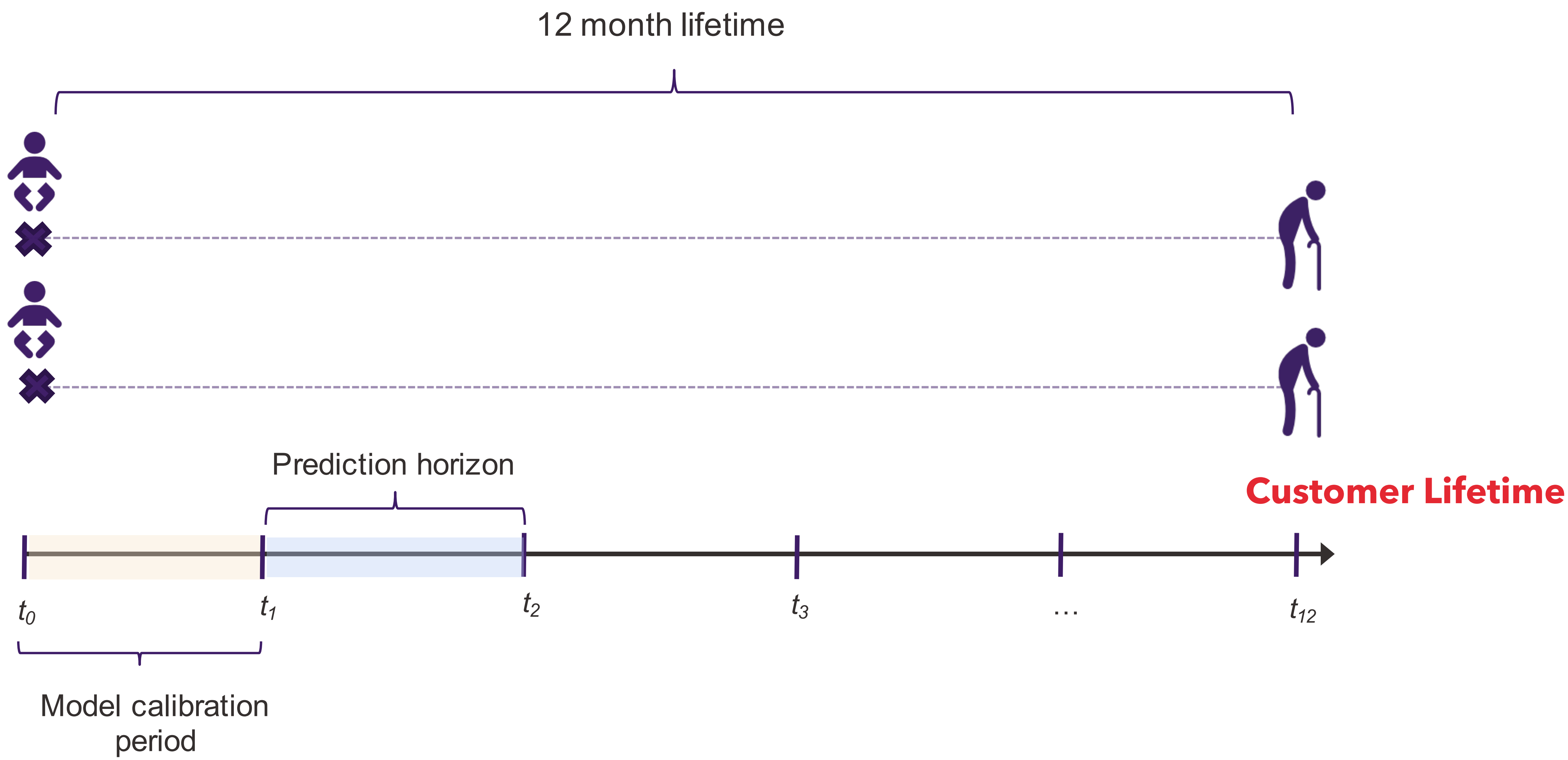
- Degree
- Page Rank
- **Cohesion**

- Avg. # of Friends of Friends
- Friends of Friends Avg. Transaction Frequency
- **FoF Cohesion**
- **Mutual Friends of Friends**

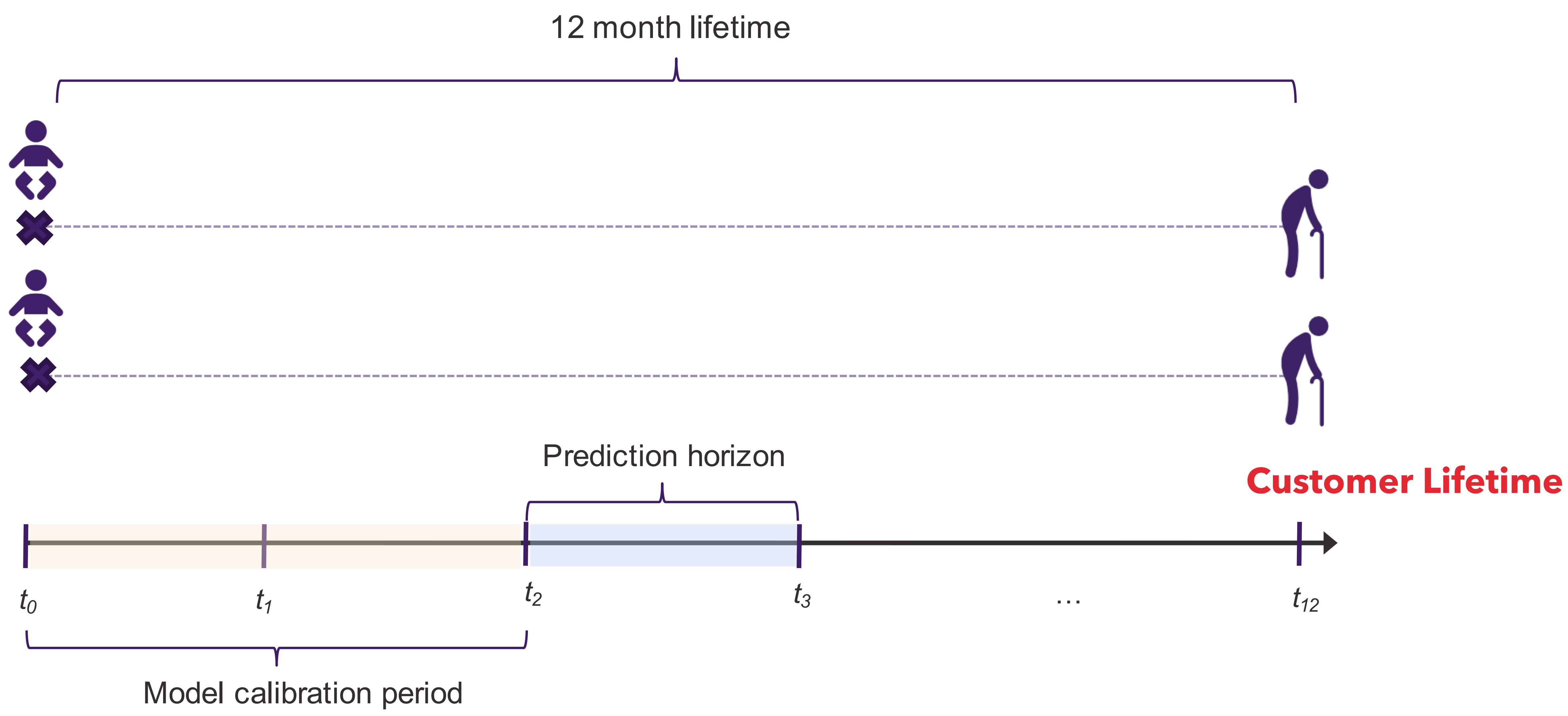


Venmo Timeframe: January 2014 - June 2016

METHODOLOGY – DYNAMIC ANALYSIS



METHODOLOGY – DYNAMIC ANALYSIS



COMPUTATIONAL REQUIREMENTS

- Size of the Data: **2 Terabytes**
- Amazon Web Services (AWS)
- PySpark

MODELS & DATASETS

Predictive Task	Problem Type	Prediction
Customer Activity	Binary classification	Active next month
Transaction Volume	Regression	# Transactions at the end of the year
Top 10% Most Frequent Customers	Binary classification	Top 10% or not

- Inputs/Datasets:
  1. User
  2. Social
  3. Both

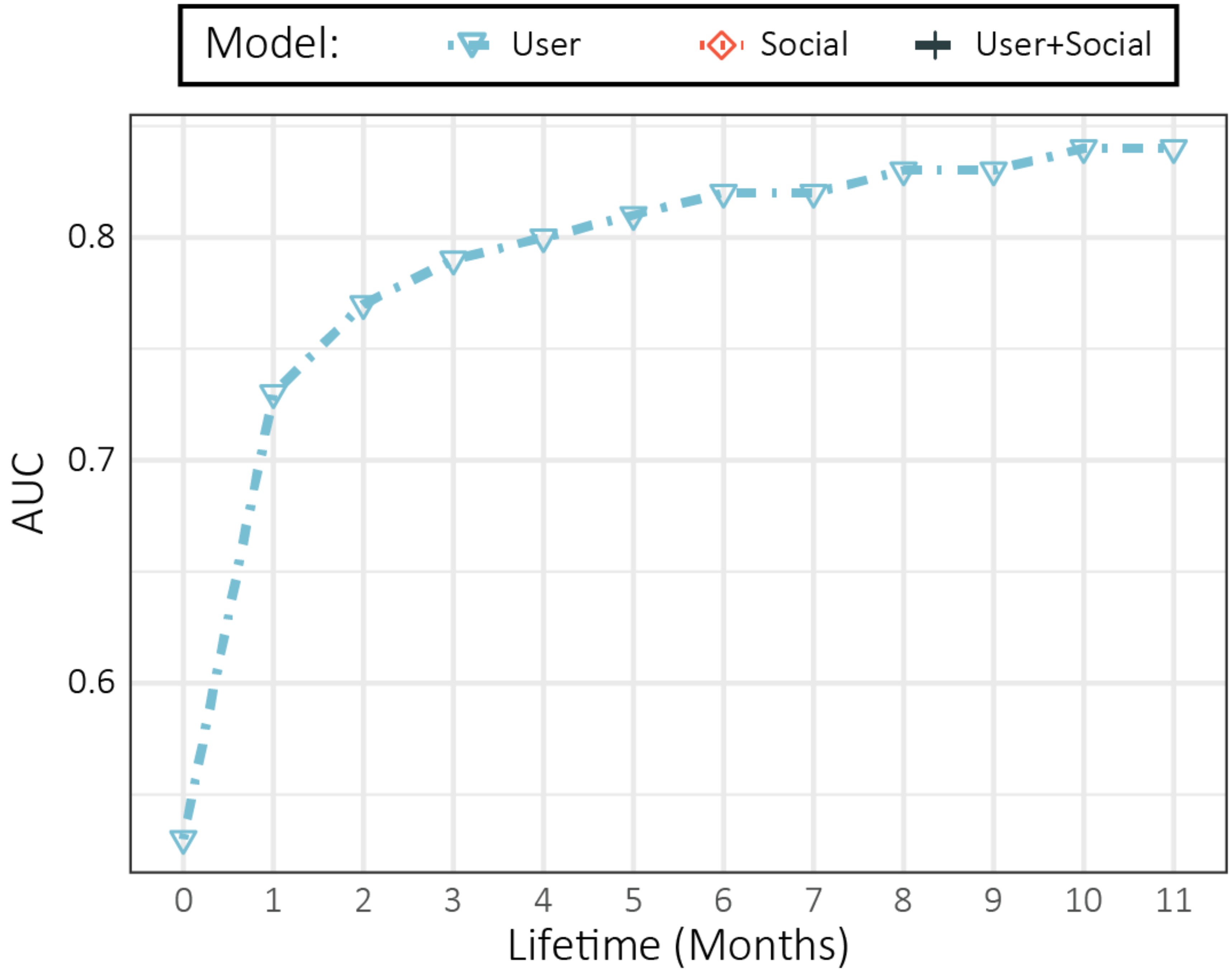
CUSTOMER ACTIVITY MODEL

- Prediction: Active in Period t+1 (0/1)

$$y_{i,t+1} = F(x_{i,t})$$

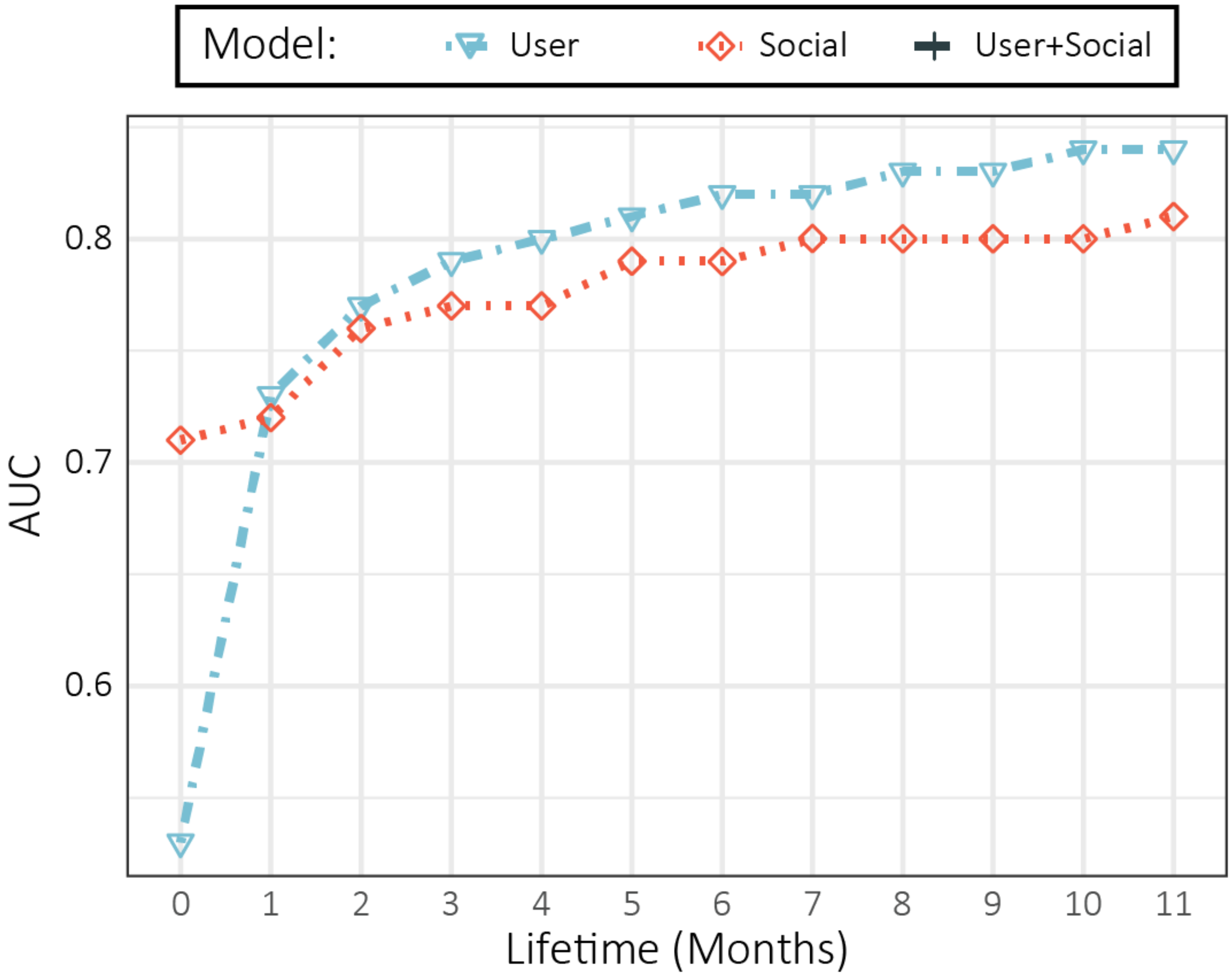
- Inputs/Data:
  1. User
  2. Social
  3. Both
- Metric/Fit: Area Under the Curve (AUC)

PREDICTING AT TIME 0 IS DIFFICULT



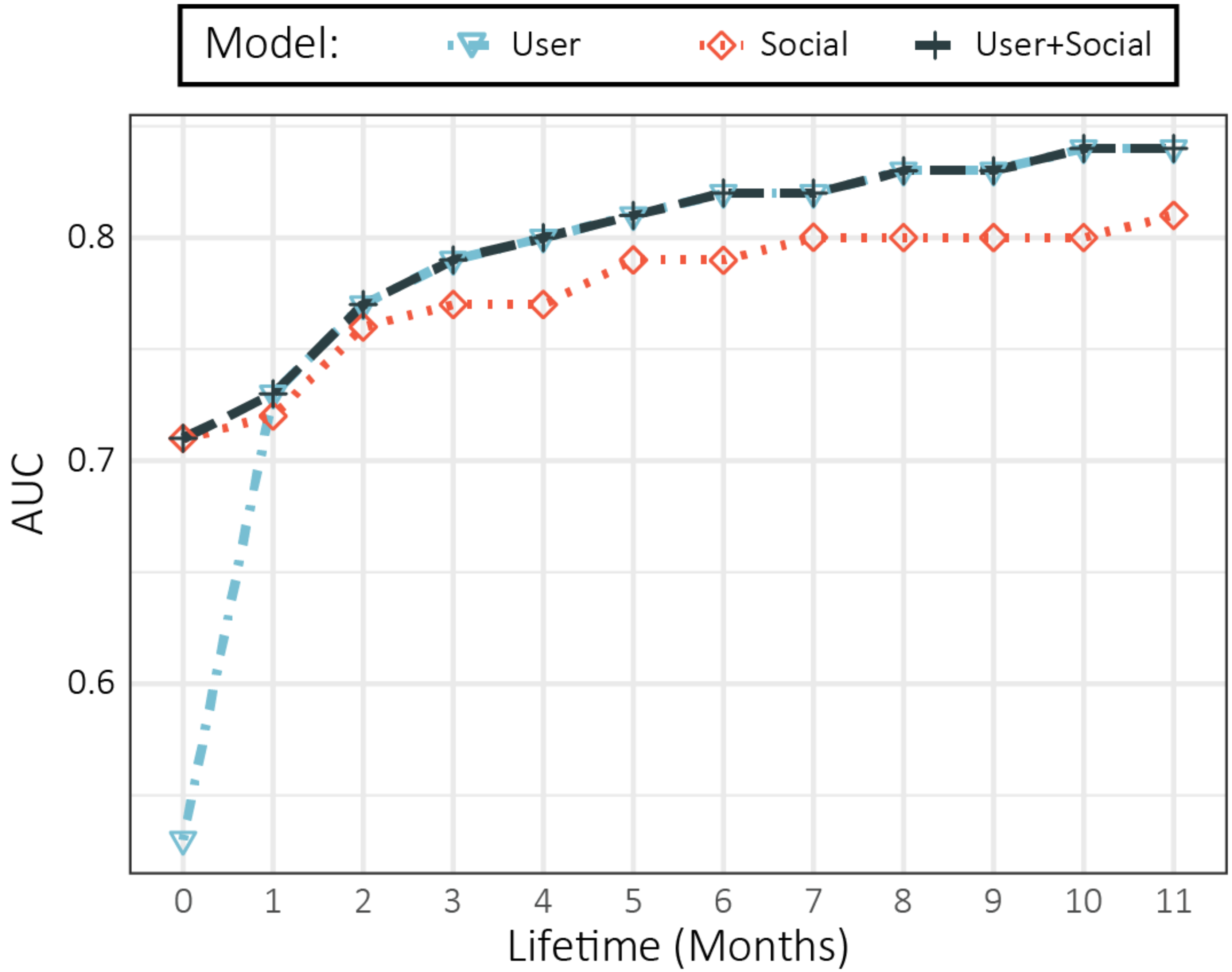
Time	User	Social
T = 0	53%	

SOCIAL NETWORK DATA IS HIGHLY PREDICTIVE AT TIME 0



Time	User	Social
T = 0	53%	**71%**

SOCIAL DATA HELPS SOLVE THE COLD-START PROBLEM BUT BECOMES IRRELEVANT AFTER MONTH 3



Time	User	Social
T = 0	53%	**71%**
T = 6	**84%**	79%

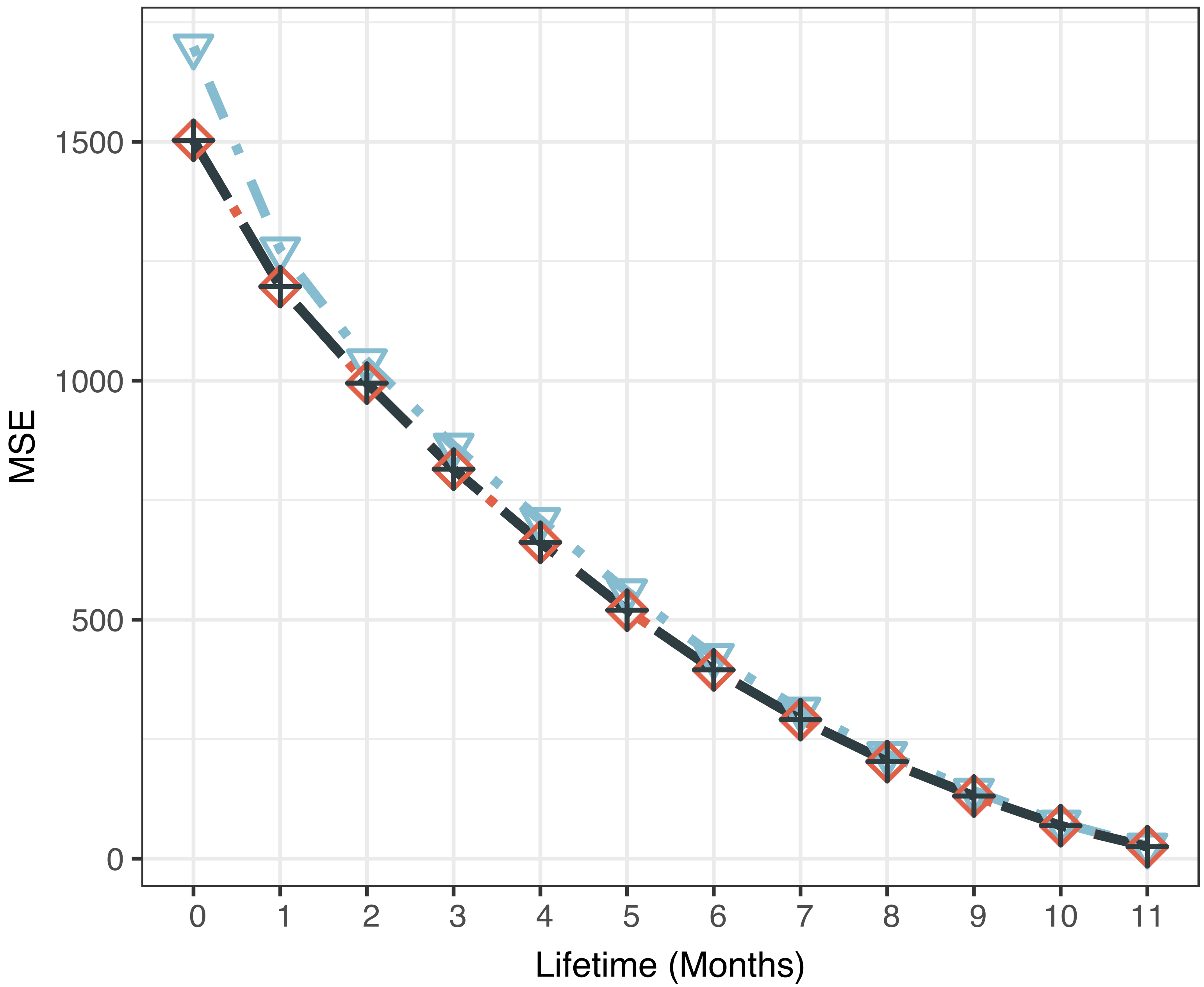
TRANSACTION VOLUME MODEL

- Prediction: Total Number of Transactions at Period 12

$$y_{i,12} = F(x_{i,t})$$

- Inputs/Data:
  1. User
  2. Social
  3. Both
- Metric/Fit: Mean Squared Error (MSE)

LONG TERM TRANSACTION VOLUME



- Cold start results also hold here
  - **Relative** Decrease at time **0**: **13% MSE**
- **Social always** achieve maximum predictive performance

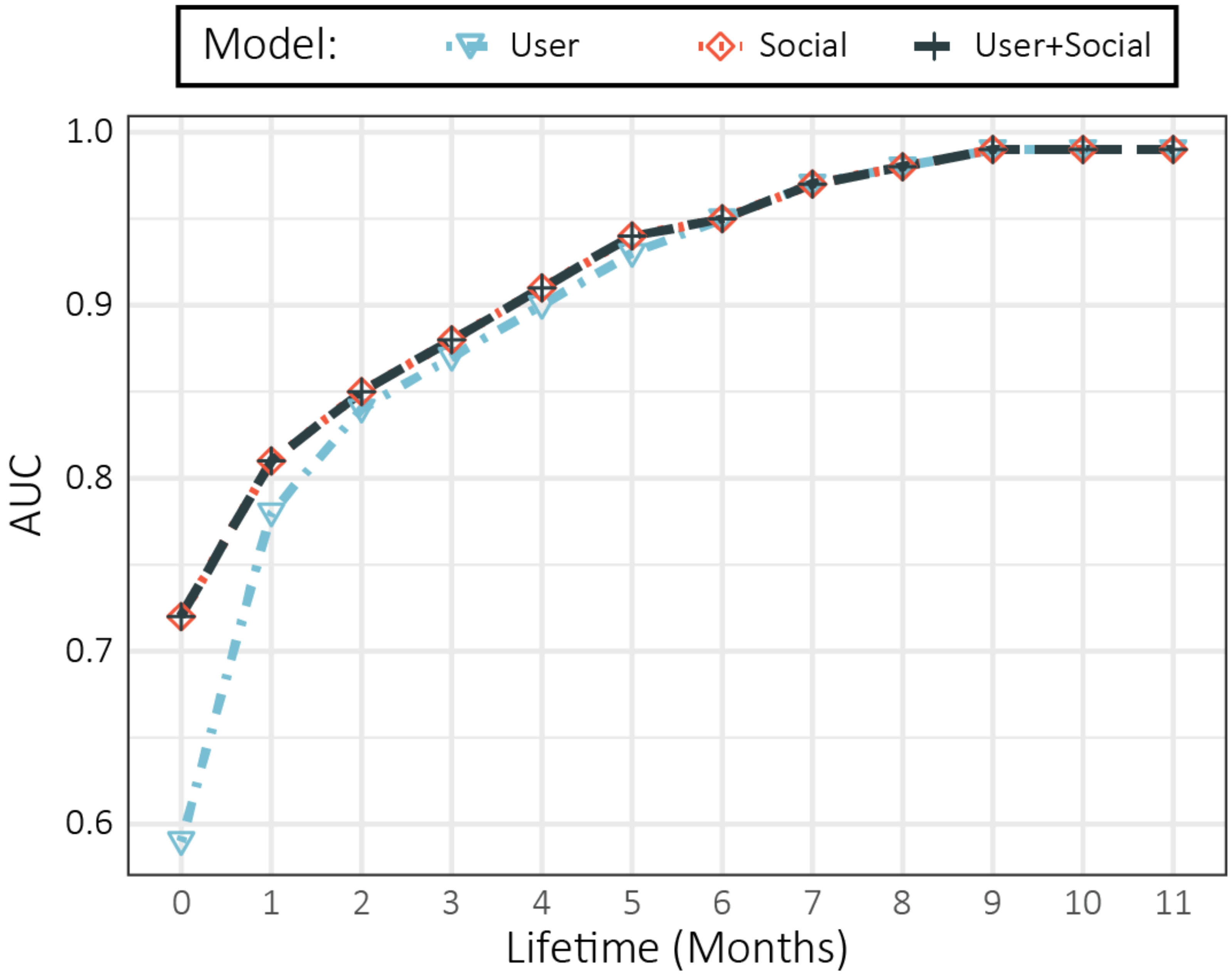
10% MOST FREQUENT CUSTOMERS MODEL

- Prediction: 10% Most Frequent Customers at Period 12

$$y_{i,12} = F(x_{i,t})$$

- Inputs/Data:
    1. User
    2. Social
    3. Both
- Metric/Fit: Area Under the Curve (AUC)

10% MOST FREQUENT CUSTOMERS



Time	User	Social
T = 0	59%	**72%**
T = 6	95%	95%

10% VS 90% MOST FREQUENT CUSTOMERS AT TIME 0

Structural Network Variables
Friends of friend's Cohesion
Mutual Friends of Friends
Outgoing Transaction %
Friend's number of friends
Friends of friend's average number of friends
Giant Component
Page Rank

10% VS 90% MOST FREQUENT CUSTOMERS AT TIME 0

Network Variables	Top 10%	Bottom 90%	Cohen's Effect Size
Friends of friend's Cohesion	0.34 (0.34)	0.30 (0.33)	0.12*
Mutual Friends of Friends	0.06 (0.14)	0.02 (0.09)	0.44**
Friends of friend's average number of friends	10.06 (7.98)	8.19 (7.69)	0.24**
Page Rank	0.32 (0.40)	0.26 (0.29)	0.22**
Giant Component	0.87 (0.33)	0.80 (0.40)	0.18*

- Most frequent customers have a more connected and denser friends of friends network whey they join the service.

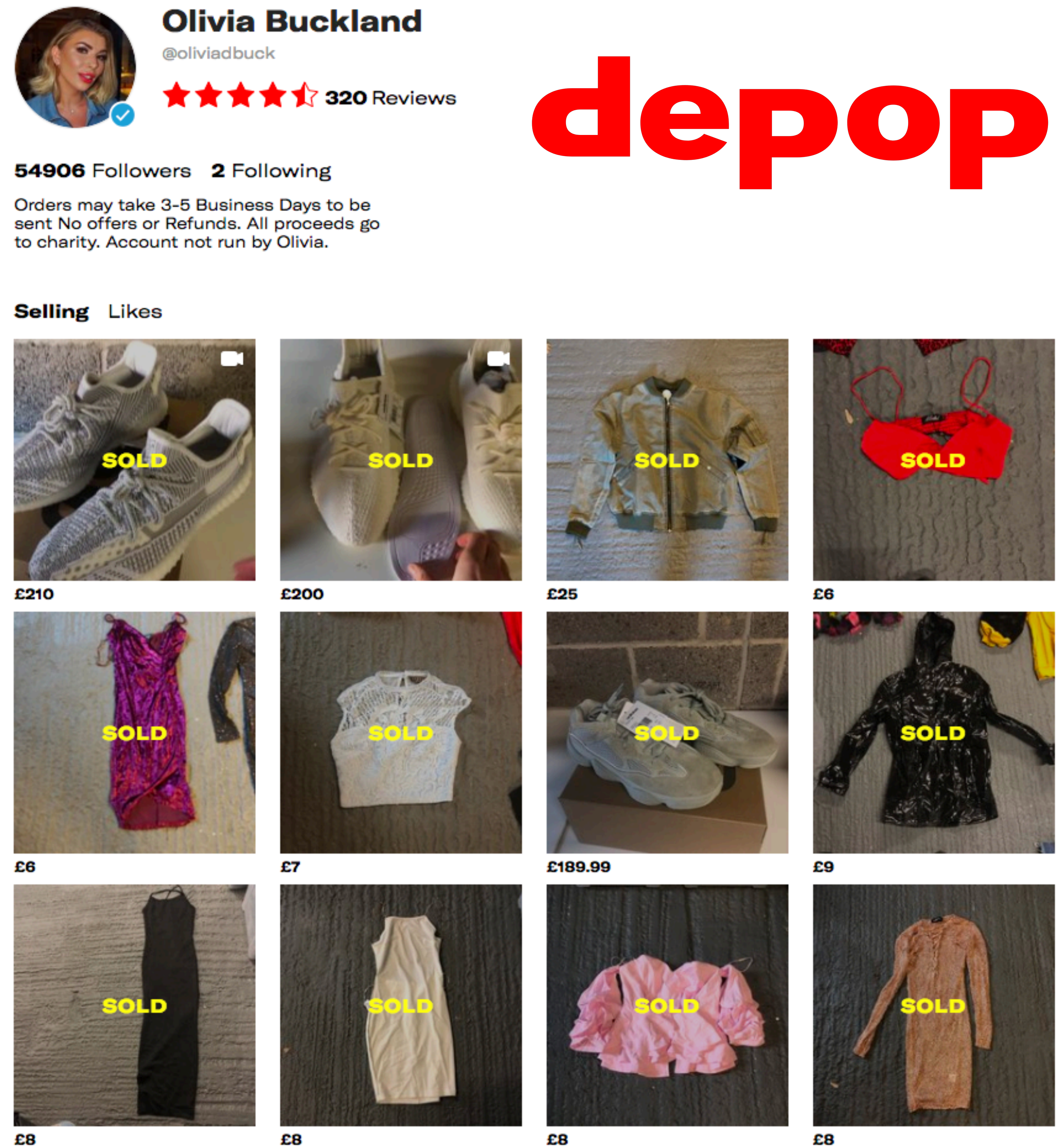
MANAGERIAL IMPLICATIONS

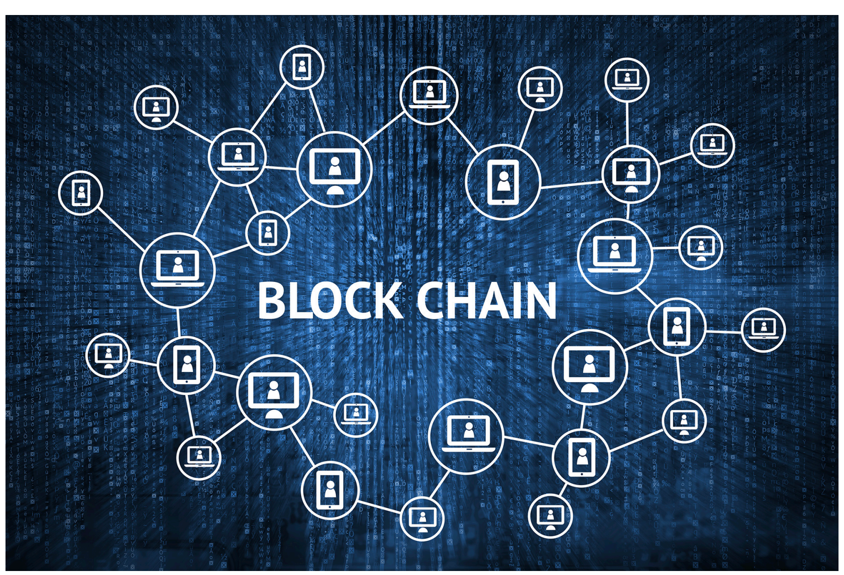
1. Marketing Resource Allocation
2. Customer Based Corporate Valuations (CBCV)

CONCLUSIONS

1. Social network data can be incredibly important when you don't have transactional data.
2. **Boundary Conditions** for when user data supplant social data:
   - True for short term activity
   - Not for long term transaction volume and top 10%
3. **Social Data Privacy Issues:** When you data are not truly yours

FUTURE RESEARCH: P2P PLATFORMS





THANK YOU VERY MUCH!

- Any Questions?
- You can always reach out to me at ploupos@ucdavis.edu
- @louposp

APPENDIX: ROBUSTNESS CHECKS

- Different ML Models

Lifetime	Model 1			Model 2			Model 3		
	Logistic	Lasso	RF	Logistic	Lasso	RF	Logistic	Lasso	RF
0	0.53	0.53	0.51	0.71	0.71	0.72	0.71	0.71	0.72
1	0.73	0.73	0.72	0.72	0.72	0.72	0.73	0.73	0.72
2	0.77	0.77	0.76	0.76	0.76	0.73	0.77	0.77	0.76
3	0.79	0.79	0.78	0.77	0.77	0.75	0.79	0.79	0.78
4	0.80	0.80	0.79	0.77	0.77	0.75	0.80	0.80	0.79
5	0.81	0.81	0.80	0.79	0.79	0.76	0.81	0.81	0.80
6	0.82	0.82	0.80	0.79	0.79	0.76	0.82	0.82	0.80
7	0.82	0.82	0.80	0.80	0.80	0.79	0.82	0.82	0.80
8	0.82	0.83	0.80	0.80	0.80	0.79	0.82	0.83	0.80
9	0.82	0.83	0.80	0.80	0.80	0.79	0.83	0.83	0.80
10	0.83	0.84	0.80	0.80	0.80	0.79	0.83	0.84	0.80
11	0.83	0.84	0.81	0.81	0.81	0.79	0.83	0.84	0.81

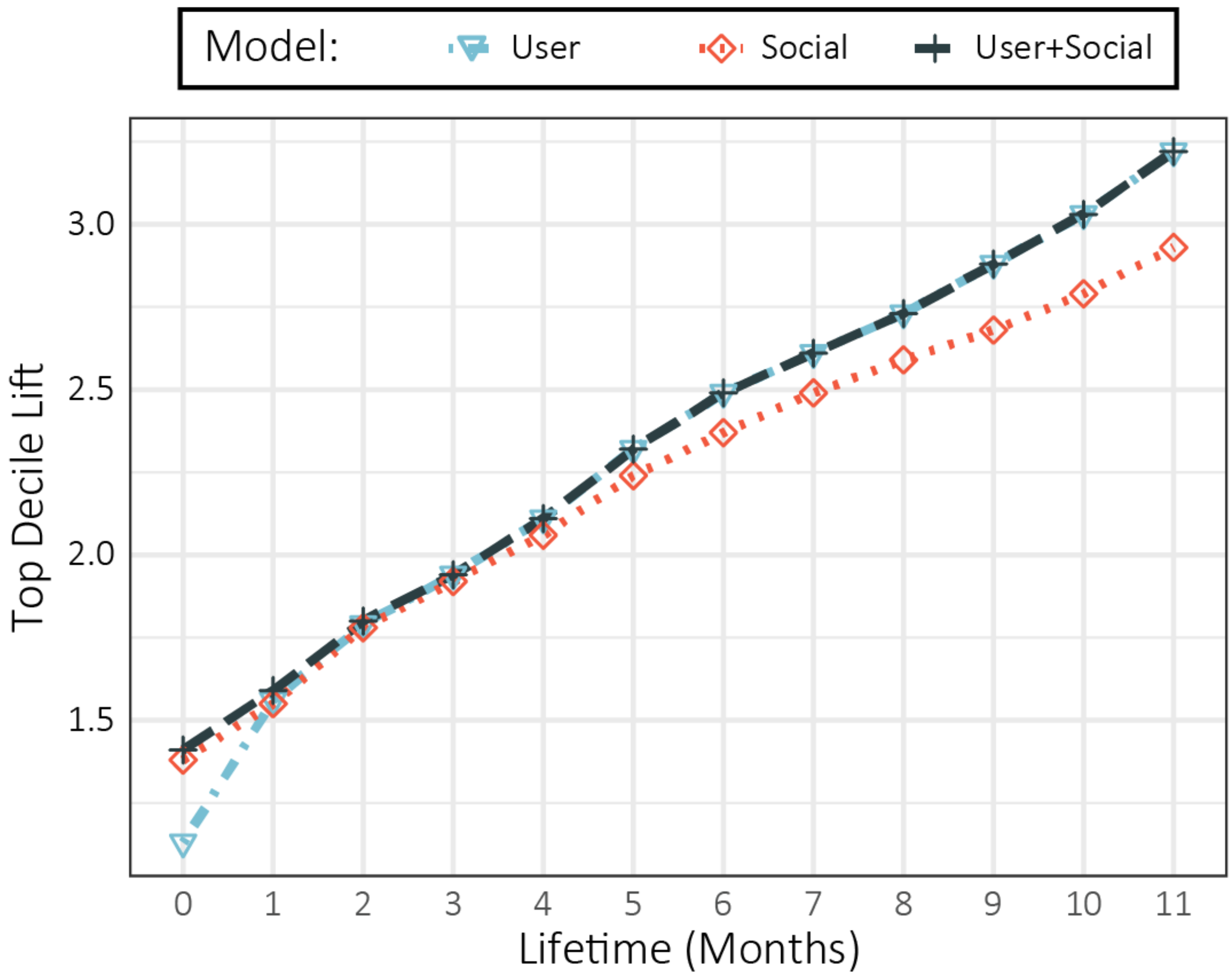
APPENDIX: ROBUSTNESS CHECKS

- Different Predictive Windows

Lifetime	60 days			90 days			120 days		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
0	0.53	0.68	0.68	0.53	0.67	0.68	0.54	0.67	0.67
1	0.73	0.72	0.73	0.73	0.72	0.73	0.73	0.73	0.74
2	0.77	0.76	0.78	0.78	0.76	0.78	0.78	0.77	0.79
3	0.79	0.77	0.79	0.80	0.78	0.80	0.81	0.79	0.81
4	0.81	0.79	0.81	0.82	0.80	0.82	0.83	0.81	0.83
5	0.82	0.80	0.83	0.83	0.81	0.84	0.84	0.82	0.84
6	0.84	0.81	0.84	0.85	0.82	0.85	0.85	0.83	0.86
7	0.84	0.82	0.84	0.85	0.83	0.86	0.86	0.83	0.86
8	0.85	0.82	0.85	0.86	0.83	0.86	0.87	0.84	0.87
9	0.85	0.83	0.85	0.87	0.84	0.87	-	-	-
10	0.86	0.83	0.86	-	-	-	-	-	-

APPENDIX: ROBUSTNESS CHECKS

- Top Decile Lift Metric



APPENDIX

- Conditional Expected # of Transactions at month 12