UNDERSTANDING BEHAVIORS THAT LEAD TO PURCHASING:

A CASE STUDY OF PINTEREST

CAROLINE LO, DAN FRANKOWSKI, JURE LESKOVEC

Team: SSHOP_249

Team Details: Team SSHOP_249

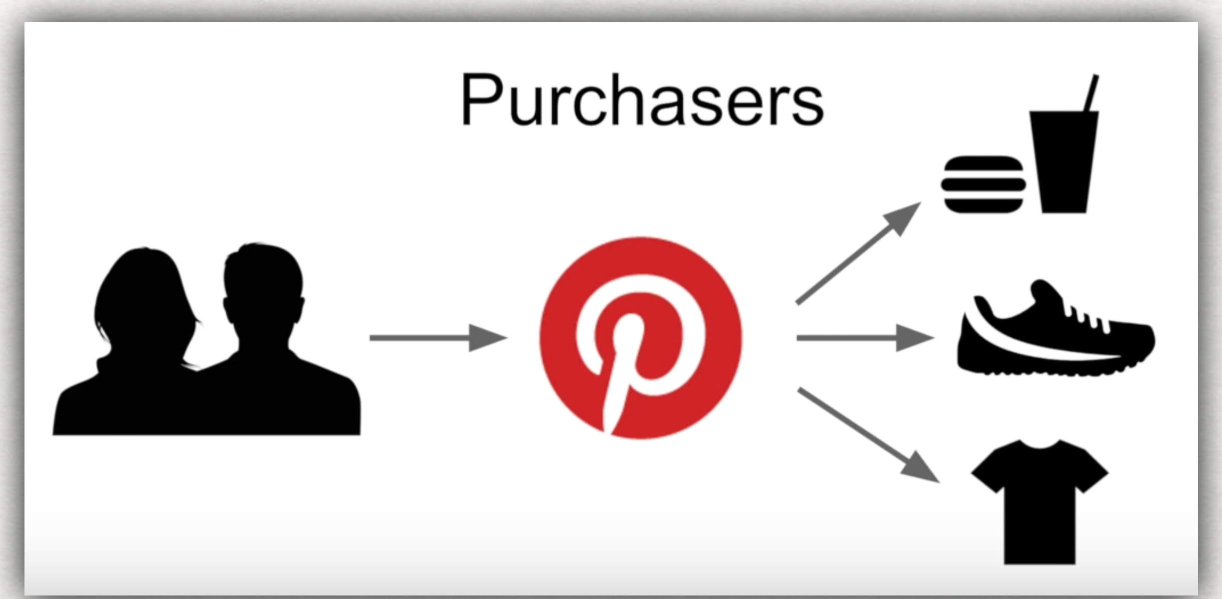
Heenal Doshi	004758927	doshi.heenal@gmail.com
Omkar Patil	904760474	omkar.9194@gmail.com
Pranav Sodhani	804591764	sodhanipranav@gmail.com
Shubham Mittal	104774903	mitshubh@gmail.com
Swati Arora	404758379	swati3124@gmail.com

AGENDA

1. Problem Statement
2. Motivation
3. Related Work
4. Dataset
5. Observations
6. Modeling Purchase Intent
7. Conclusion
8. Future Work

PROBLEM STATEMENT

- Understand and Model long term, time varying user purchasing intent using (CDA) Content Discovery Applications.



WHY?

- Online shopping in US alone - 350 bn USD/ year.

    + growing at ~15% annually.

- Improved advertising, search results ranking and personalization

    eg. 1. promote buyable pins to users identified with purchase intent.

    2. target pricing

RELATED WORK

- Collaborative Filtering: Exploit other users' purchase histories
- Content based
  - product reviews, demographic data, previous purchase history
- Using short-term user activity - whether a given user session will result into purchase
  - Task Completion, Hidden Markov Models
- Using external sources
  - social media, email-data, social networks
- Used in combination

DATASET

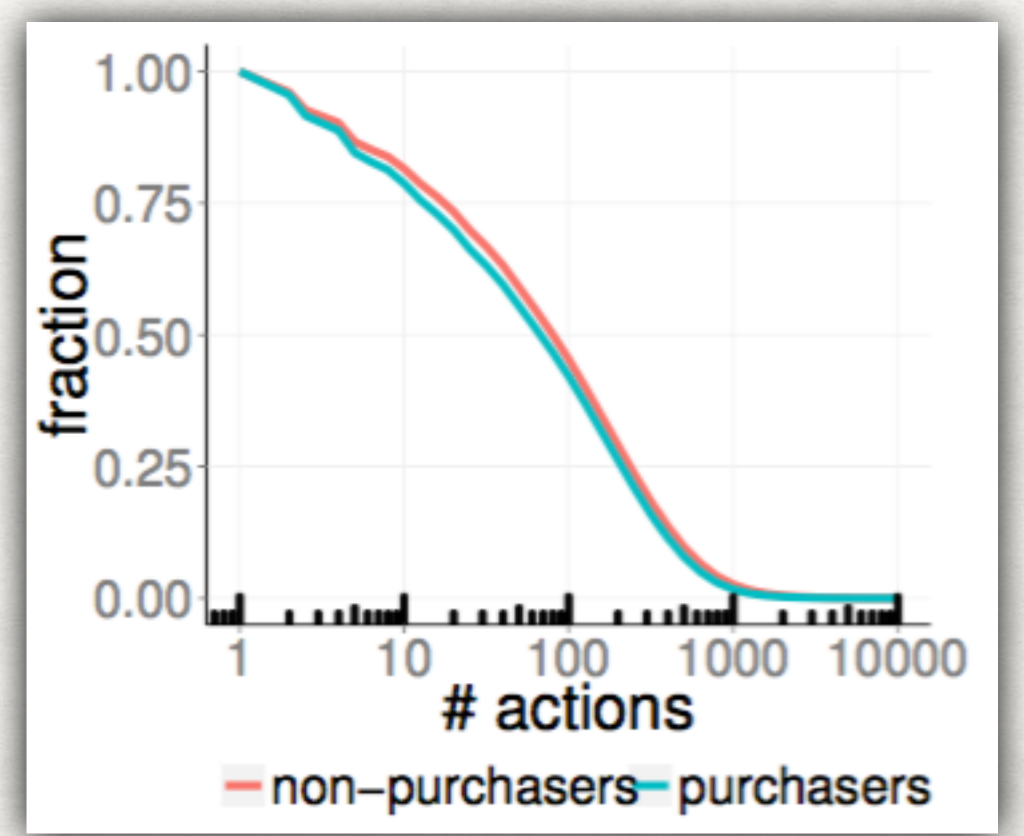
- Content Discovery Application - 4 classes of actions
  - Search
  - Closeup (Engage on-site)
  - Clickthrough (Engage off-site)
  - Save

- Pinterest Dataset

Statistic	Value
Number of purchasers	1.3M
Number of non-purchasers	1.3M
Number of users (total)	2.6M
Number of actions (total)	0.5B
Time duration	28 days
Month when purchases were made	May 2015

- <40% users make >1 purchase in May'15: randomly sampled single purchase
- Each purchaser matched with a non-purchaser by:
  - Registration month
  - #Active days before 28 days (Long term activity)
  - Gender (US based users only)

- User Activity Analysis
    - Purchasers are <u>not</u> more engaged!
    - On the surface purchasers and non-purchasers seem no different.

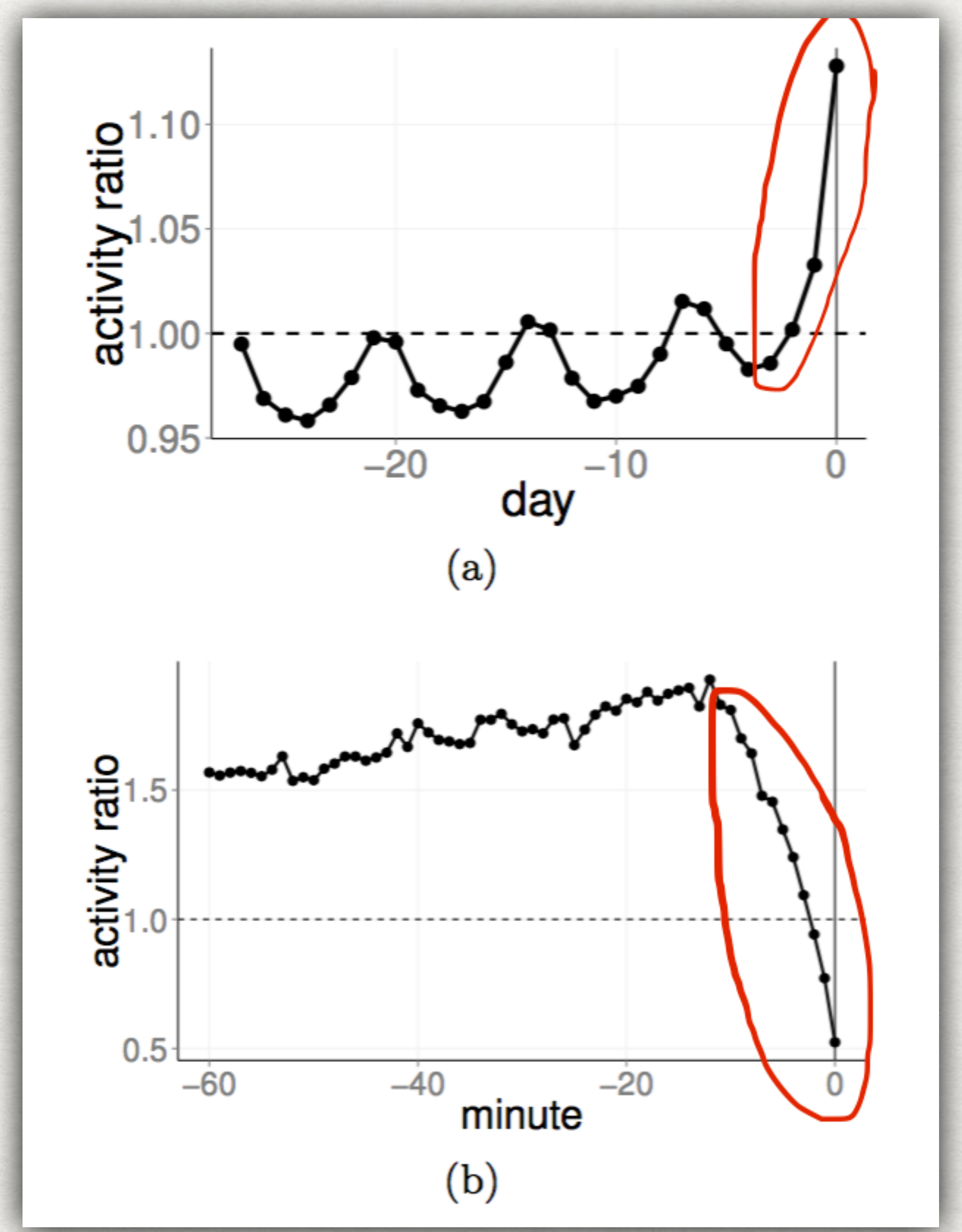


DYNAMICS OF PURCHASE INTENT

HOW DOES IT EVOLVE OVER TIME?

- Overall user activity level
  - Relative activity level of matched purchasers and non-purchasers

$$AR(d) = \frac{\%\ \text{purchasers active on day } d}{\%\ \text{non-purchasers active on day } d}.$$



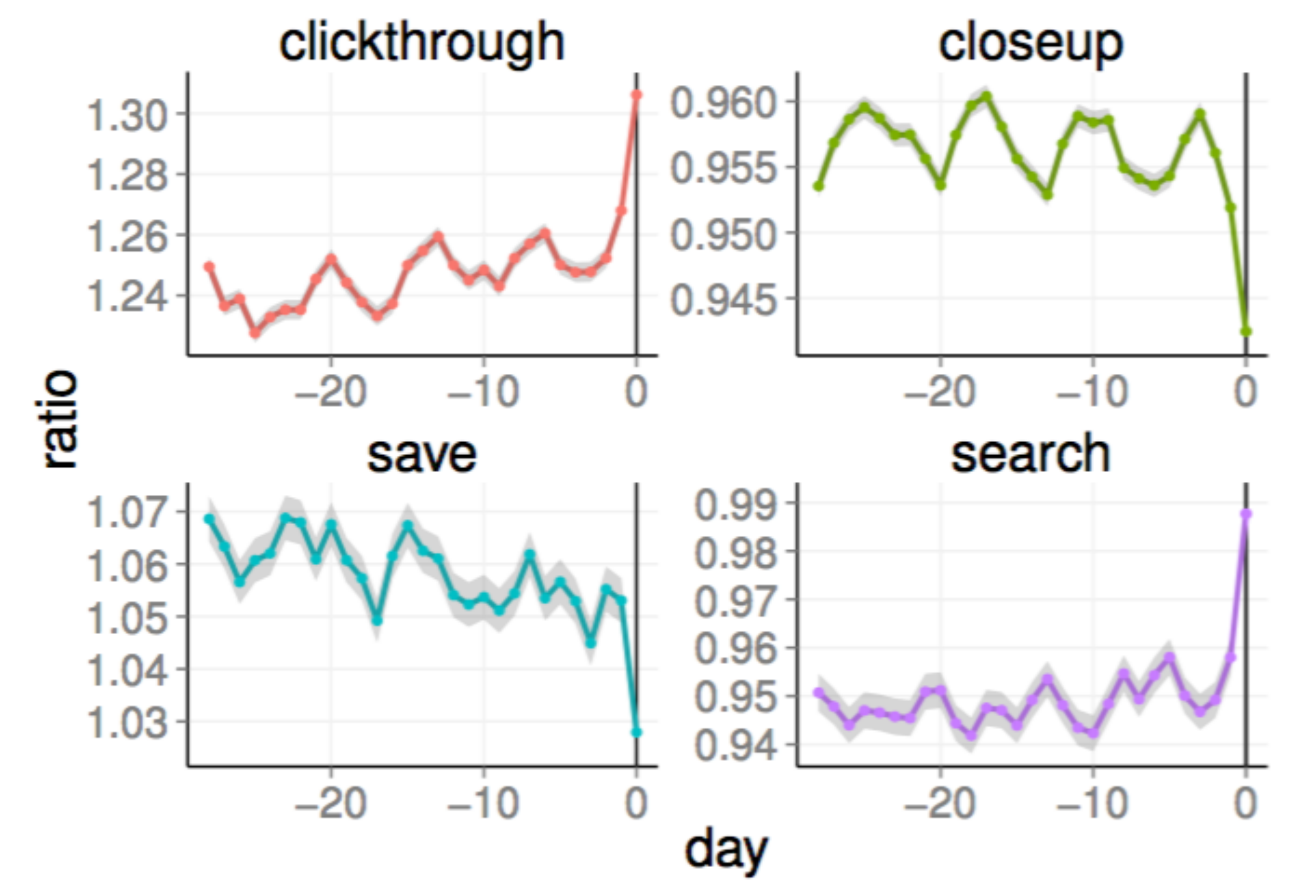
- Specific behavior - user action
  - Intentional actions
  - Action score: prioritize which action?

$$AS(u,d,a) = \frac{\text{\# actions } u \text{ took on day } d \text{ of type } a}{\text{\# actions } u \text{ took overall on day } d}.$$

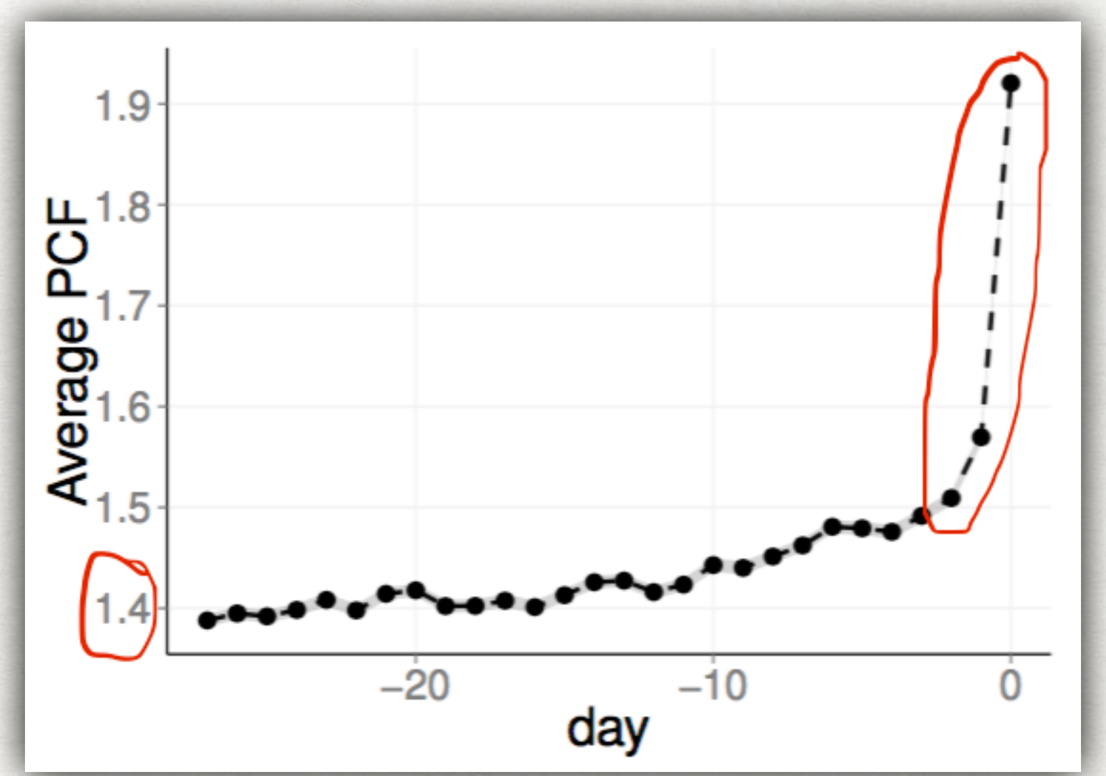
  - AS Ratio: comparing non-purchasers and purchasers.

$$ASR(d,a) = \frac{avg_{u \in \text{purchasers}} AS(u,d,a)}{avg_{u \in \text{non-purchasers}} AS(u,d,a)}.$$

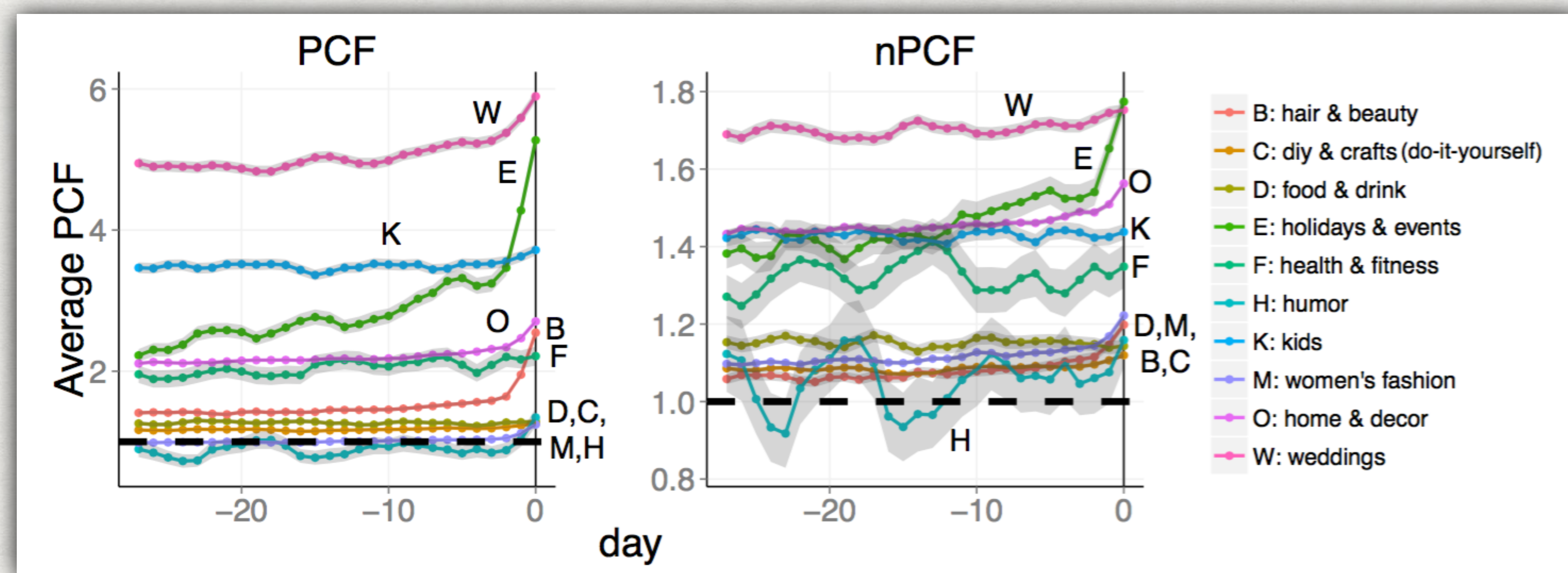
  - Purchaser behavior does change right before day 0!



- Purchasers category focus
  - Pinterest - 32 predetermined high-level categories
  - Partner companies - fall within 10 categories based on pins
  - Category interaction score (I): fraction of actions taken by a user that fall in category c on a day d.
  - Purchase category focus:



$$PCF(p,d,c) = \frac{I(p,d,c)}{avg_{u \in non-purchasers} I(u,d,c)}.$$



- PCF spike before day 0 - differs across categories (Impulsive purchase?)
- nPCF: how interested is a purchaser compared to users already interested in that category?

- Summarizing Purchase Intent Dynamics as Day 0 approaches,
  - Purchaser more likely to be active on CDA.
  - Increased investment in specific content (clickthrough and save) over a lengthy period.
  - Sharp increase in searching for and browsing through on Day 0 compared to other days.
  - Purchaser focuses on content related to their purchase and increasingly so.
  - Extent to which purchase intent is expressed varies across categories.

MODELING PURCHASE INTENT

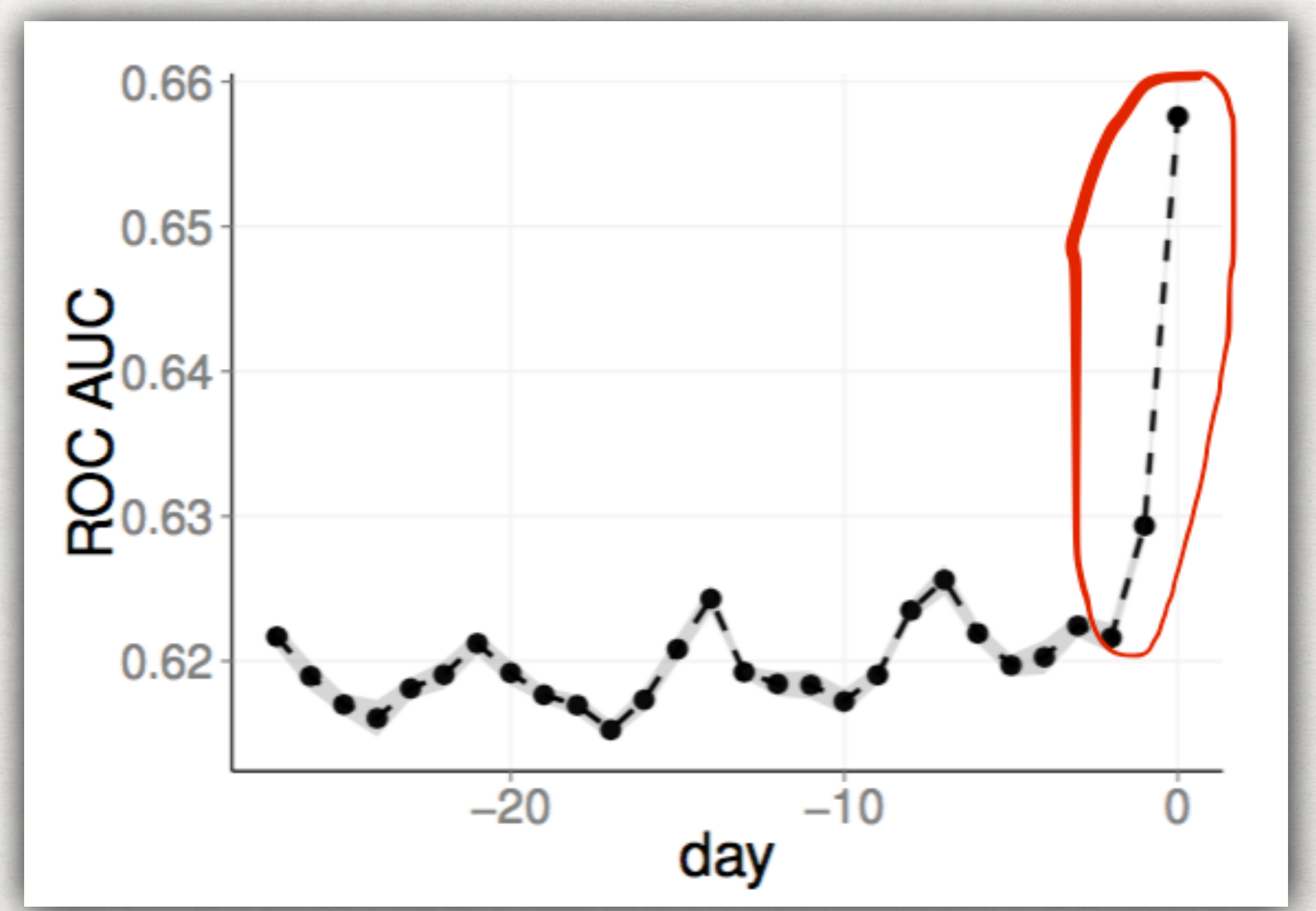
Identifying purchaser based on their behavior on day -k

- Features for learning task
  - Demographics: Gender, Region, #Boards, #Pins
  - Activity: #Active days in last 7 days, #Actions/day
  - Action-type: Action-ratio for each action type, Platform usage
  - Content: Category interaction scores (CIS)
  - Temporal: Difference between Action-type & Content features values over an interval of a) 3 days, b) 30 hours & c) 7 days prior.

- Modeling Purchase Intent Over Time
  - Task 1: How well can purchasers be identified as Day 0 approaches?
    - Sample 250K purchases & non-purchasers, active on -kth day
    - Train separate logistic regression classifiers per day k (0 to -27) and compute the 10-fold cross validated AUC ROC curve.

Observations:

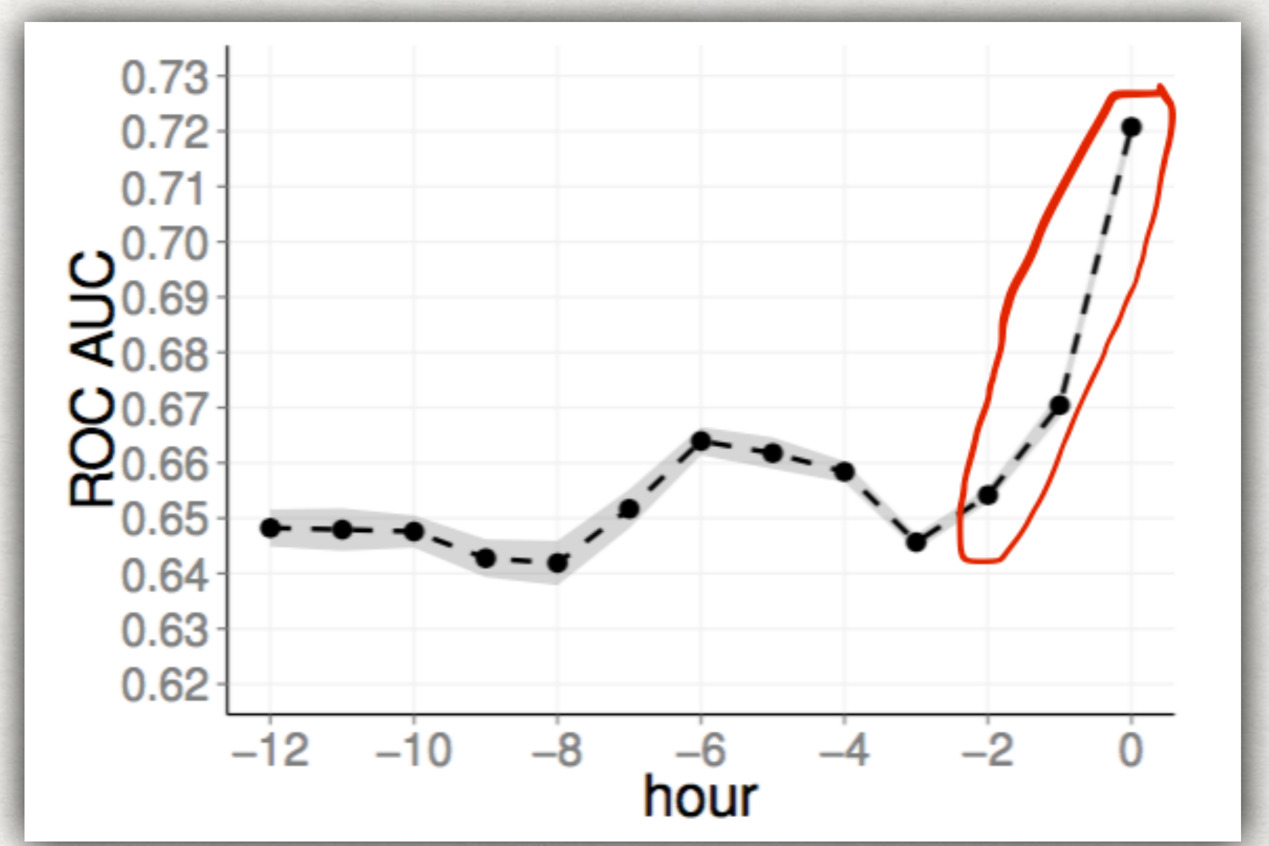
- ROC AUC ~ [0.62, 0.66]
- Purchase intent builds slowly and spikes in the last 3 days



- Task 2: Predict if a user active at -k hours is a purchaser or not?

Observations:

- ROC AUC ~ [0.65, 0.72]
- Short-term differences of purchasers exist shortly before a purchase is made



- Examining Signal Importance
  - Identifying purchasers active on day 0
  - Identifying purchasers active on hour 0

Features	**Day 0**	**Hour 0**
Random Baseline	0.500	0.500
Demographics (d)	0.513 ± 0.0007	0.513 ± 0.0024
Activity (a)	0.591 ± 0.0008	0.625 ± 0.0015
Action-type (A)	0.633 ± 0.0008	0.698 ± 0.0020
Content (C)	0.577 ± 0.0008	0.617 ± 0.0020
Temporal (T)	0.544 ± 0.0007	0.581 ± 0.0019
d + a + A (*how*)	0.642 ± 0.0009	0.702 ± 0.0017
d + a + A + C + T	$\mathbf{0.657} \pm 0.0007$	$\mathbf{0.721} \pm 0.0026$

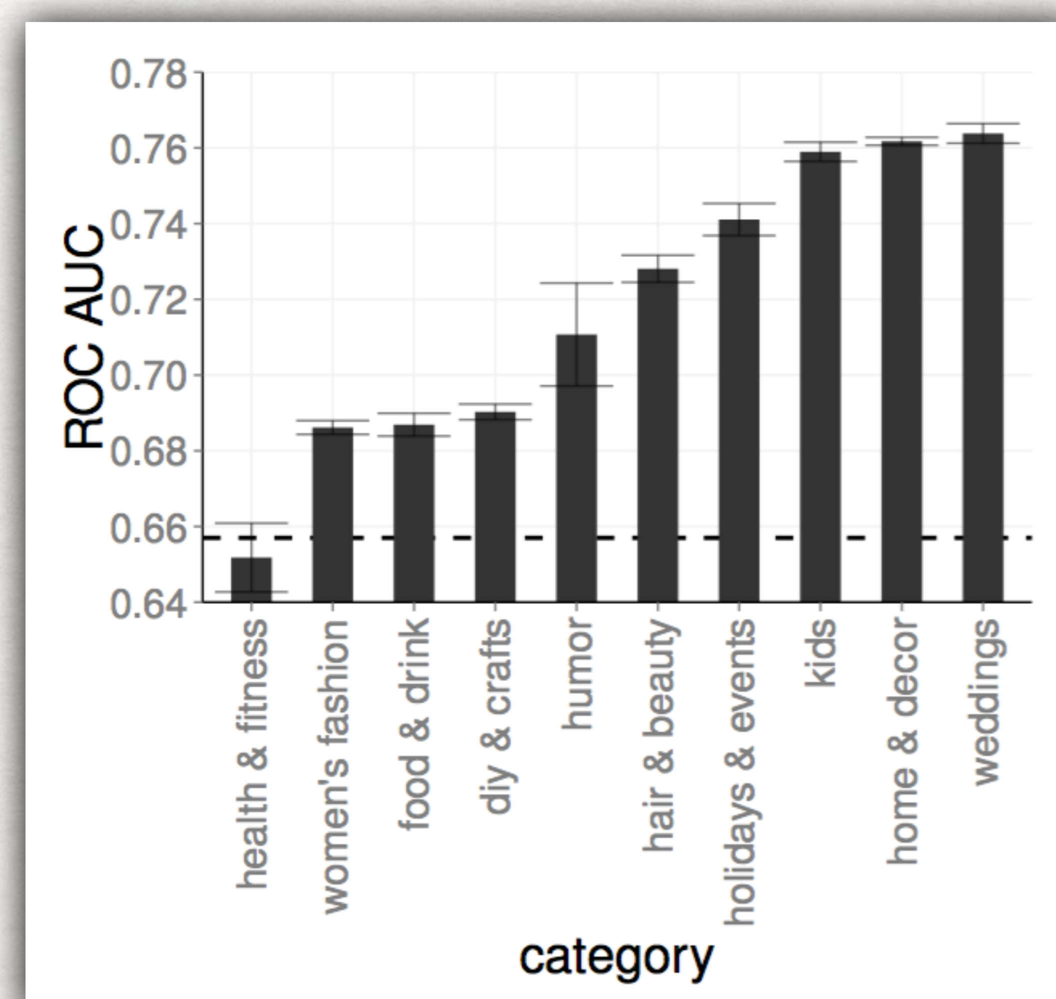
- Modeling per-category purchase intent
    - Task: Predict whether a user will purchase an item in a specific category.
    - Sample active users on the purchase date and build one binary classifier for each category

Observations:

- Differences in ROC values ⇒

  purchase intent is stronger for some categories

- Purchase intent can be modeled both over time, as well as within a specific category

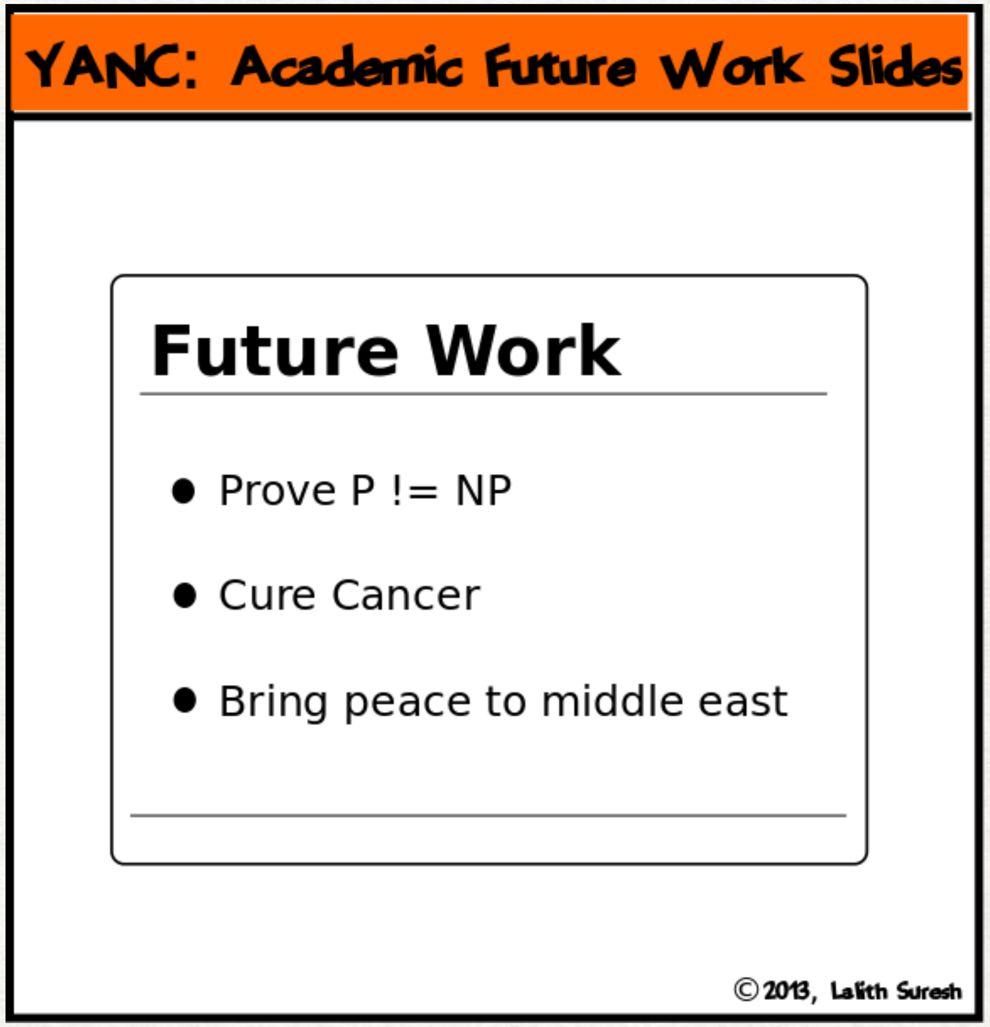


CONCLUSION

- A data-driven, cross-platform approach.
- One of its type -- longitudinal study of the dynamics of user study
- Serves as a basis for future studies on the intersection of online behavior and purchasing patterns
- Results not reproducible - performance on other datasets ?
- Mathematical Relations guided by heuristics rather than previously proven methods

FUTURE WORK

- How to handle impulsive shoppers ?
- Personalize search results based on purchasing intent
- Buyable Pins ?



THANK YOU

QUESTIONS?

BIBLIOGRAPHY

[1] Z. Huang, D. Zeng and H. Chen, "A Comparison of Collaborative-Filtering Recommendation Algorithms for E-commerce," in IEEE Intelligent Systems, vol. 22, no. 5, pp. 68-78, Sept.-Oct. 2007.doi: 10.1109/MIS.2007.4338497

[2] Jiangtao Qiu, Zhangxi Lin, and Yinghong Li. 2015. Predicting customer purchase behavior in the e-commerce context. 15, 4 (December 2015), 427-452. DOI=http://dx.doi.org/10.1007/s10660-015-9191-6

[3] Farshad Kooti, Kristina Lerman, Luca Maria Aiello, Mihajlo Grbovic, Nemanja Djuric, and Vladan Radosavljevic. 2016. Portrait of an Online Shopper: Understanding and Predicting Consumer Behavior. In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM '16). ACM, New York, NY, USA, 205-214. DOI: https://doi.org/10.1145/2835776.2835831

[4] Yongzheng Zhang and Marco Pennacchiotti. 2013. Predicting purchase behaviors from social media. In Proceedings of the 22nd international conference on World Wide Web (WWW '13). ACM, New York, NY, USA, 1521-1532. DOI: https://doi.org/10.1145/2488388.2488521

BIBLIOGRAPHY

[5] G. Linden, B. Smith and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," in IEEE Internet Computing, vol. 7, no. 1, pp. 76-80, Jan/Feb 2003. doi: 10.1109/MIC.2003.1167344

[6] Wang, Y., Li, J., Liu, Q. et al. J Supercomput (2015) 71: 3320. doi:10.1007/s11227-015-1495-8

[7] Sismeiro, C., & Bucklin, R. E. (2004). Modeling purchase behavior at an e-commerce web site: A task-completion approach. Journal of marketing research, 41(3), 306-323.