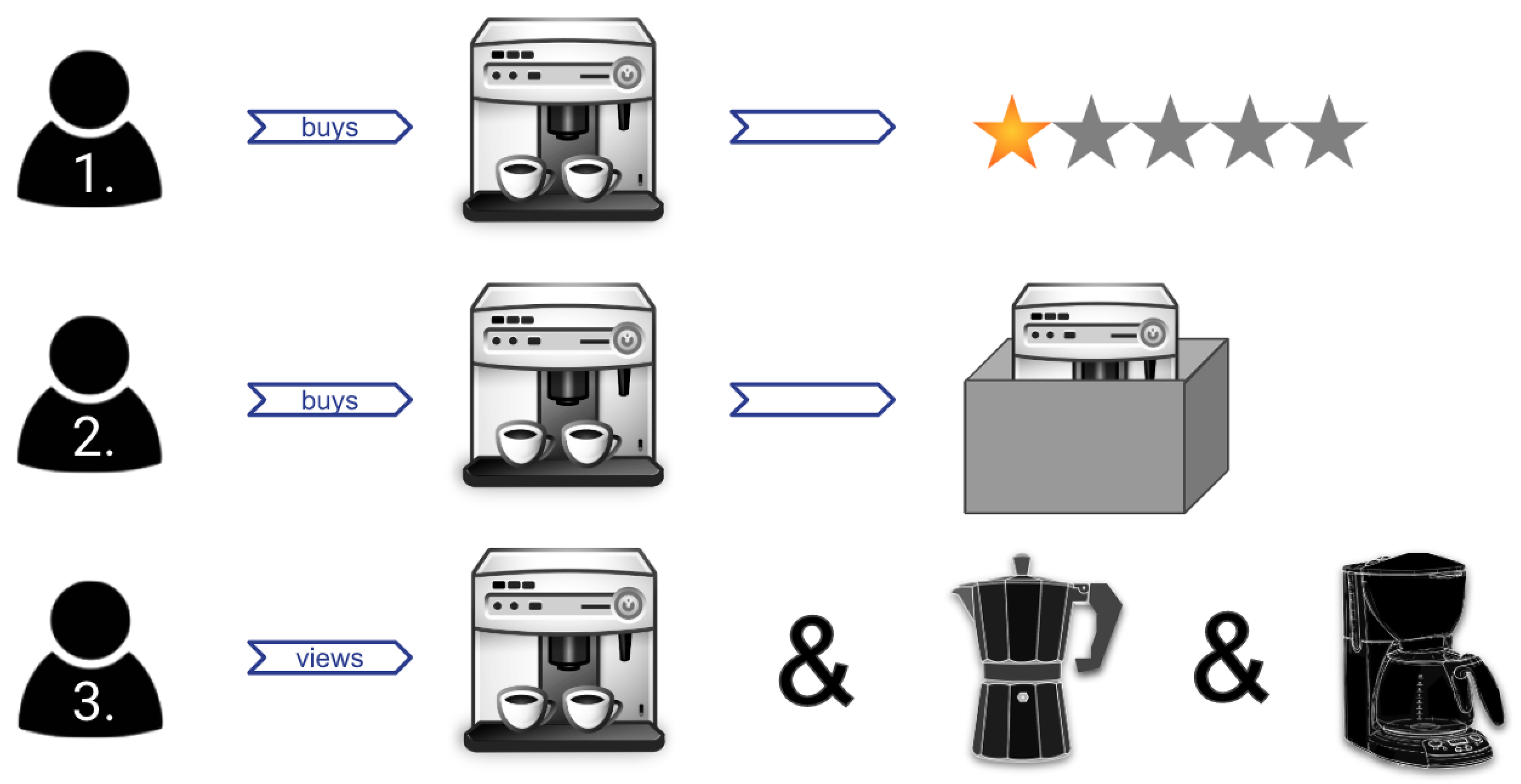
Using Embeddings of Line Graph Powers to Retrieve Item Substitutes

Dora Jambor[1,*] Brooke Fitzgerald[2,*] Putra Manggala[2,*]

[1]McGill University, RLLab, MILA [2]Shopify Inc. [*]Authors contributed equally. Work done while at Shopify, Montreal.

Motivation

In real-world recommender systems, differentiating between substitutes and complement products is important in order to capture users' intent.



Our research question in this work is the following:

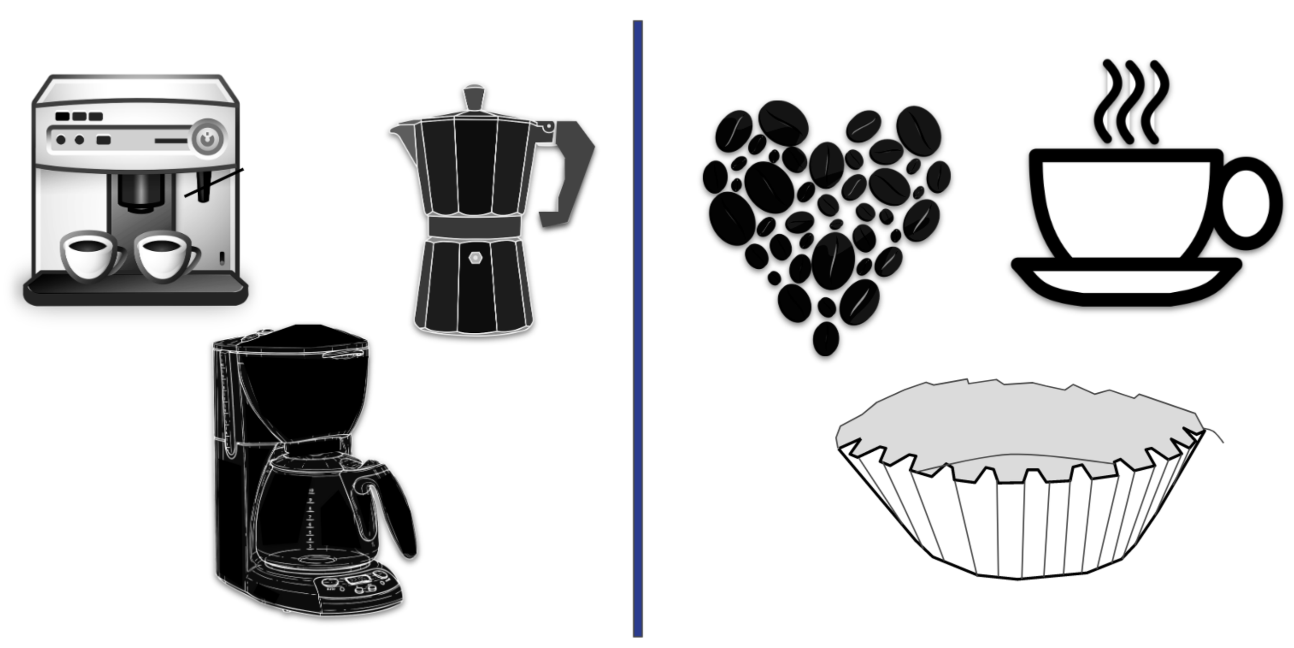
◇ Can we retrieve item substitutes by only leveraging structural information from users' purchase graph?

Complements & Substitutes

Consumer choice theory [1] defines compliments and substitutes as follows:

◇ **Complements:** Products used in conjunction with other products, often co-purchased by users.

◇ **Substitues:** Products often perceived as the same or comparable. That means, having one product would make users desire the other substitutes less.



Previous Approaches

◇ Item-based collaborative filtering on users' navigation logs: heuristics are used to map navigation logs to either substitute or complement [2]

◇ Link prediction with supervised language modeling from users' navigation logs and item meta data [3]

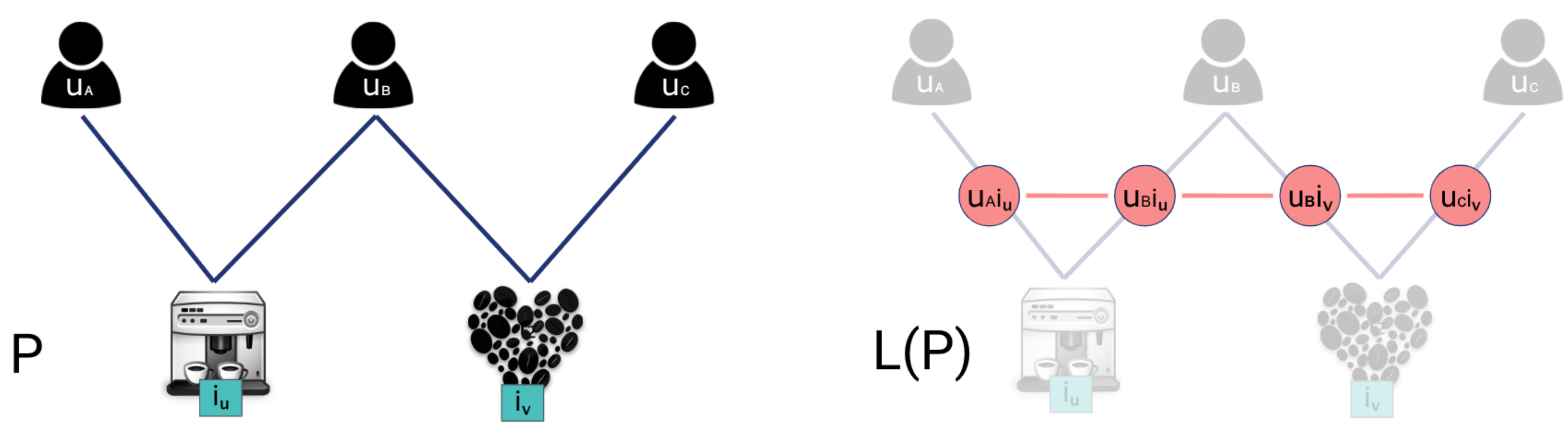
◇ Node2Vec is a random-walk based method that learns node embeddings that can be used to retrieve relevant items, however it cannot identify substitutes [4]

Setup

Let U be the set of users and I the set of items, and $P = (U \cup I, E)$ a bipartite purchase graph, with edges (u, i), for $u \in U$, and $i \in I$ for each purchase of item i by user u.

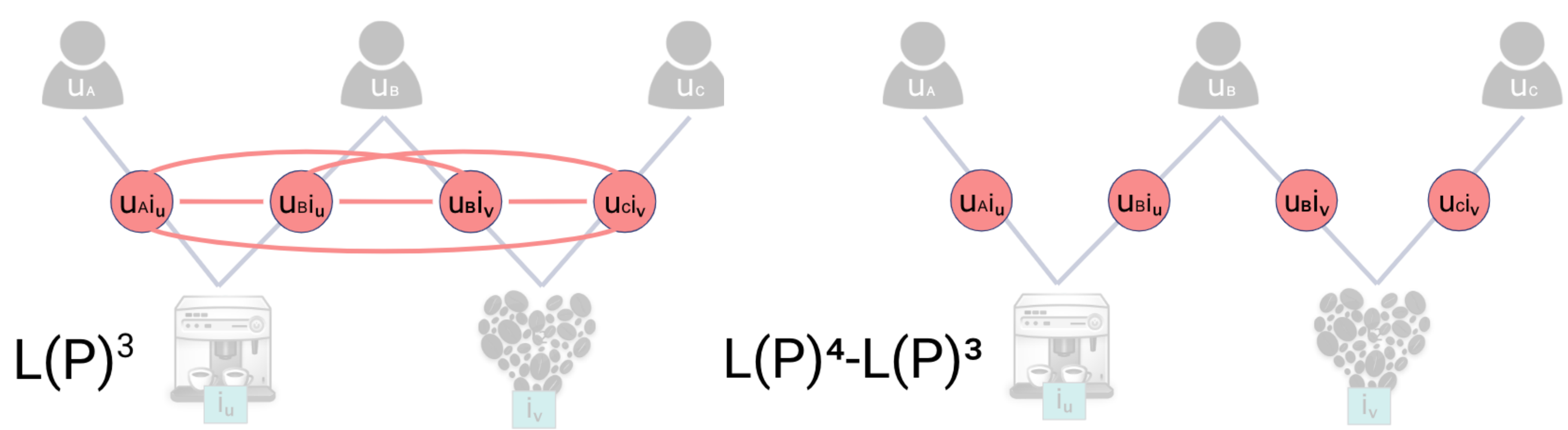
Methodology

Our method is based on the assumption that generally users co-purchase complements, not substitutes. Let L(P) denote the line graph of P, with nodes representing purchases by (u, i) pairs, and edges representing a shared item, or user between two nodes.



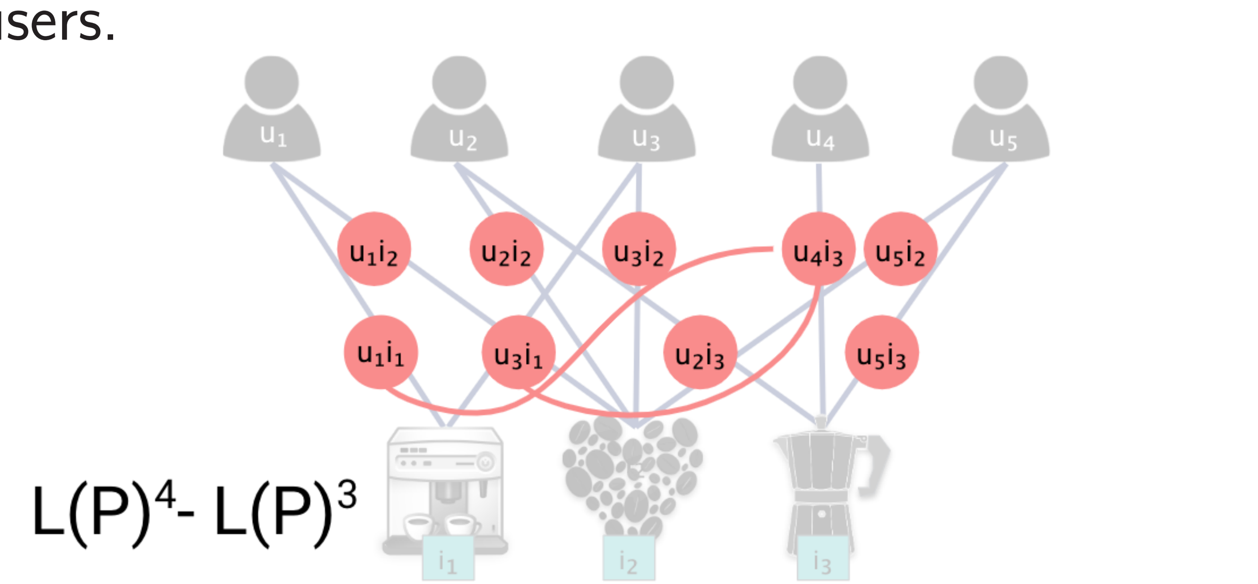
Complements Line graph of P

Let us know define $L(P)^k$, the k^{th} power of L(P), a new graph with the same set of nodes, where two nodes are adjacent if their distance is at most k. $L(P)^3$ then captures all complements.



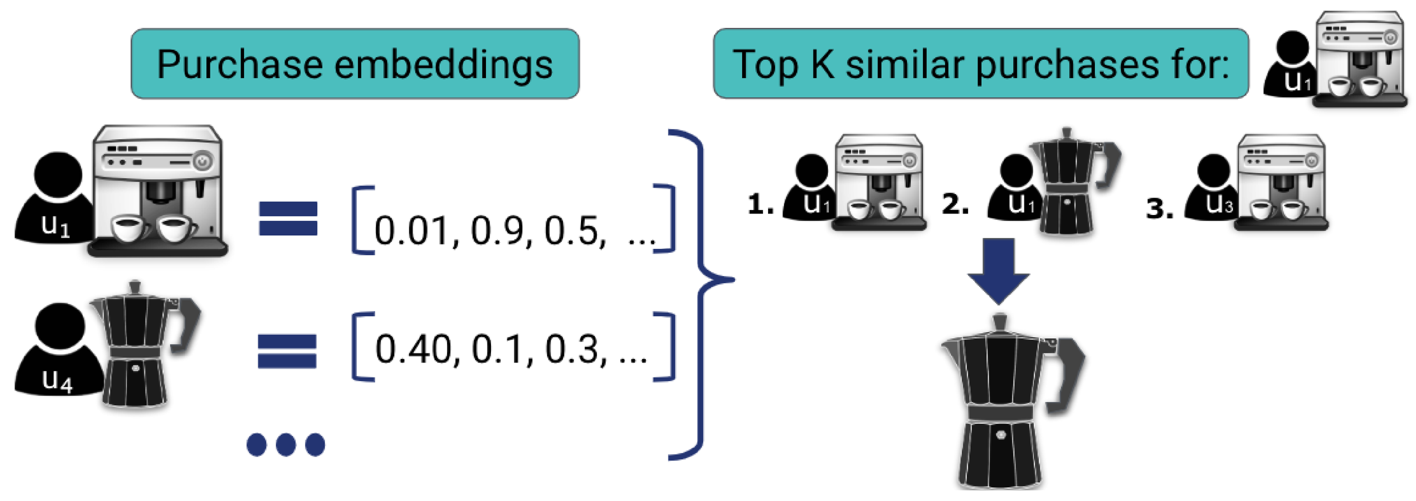
Complements Remove complement edges

$L(P)^4$ - $L(P)^3$ then defines a graph where any two nodes (i.e. purchases) have items that are never in the same purchase set of any users.



Nodes are adjacent when purchases include substitute items

Random-walk based techniques like [4] can help us generate node embeddings for $L(P)^4$ - $L(P)^3$ where substitutes are now closer in vector space.



Bibliography

[1] Mas-Colell, A., Whinston, M. D., Green, J. R. (1995). Microeconomic theory (Vol. 1). New York: Oxford university press.

[2] Zheng, J., Wu, X., Niu, J., Bolivar, A. (2009, July). Substitutes or complements: another step forward in recommendations. In Proceedings of the 10th ACM conference on Electronic commerce (pp. 139-146). ACM.

[3] McAuley, J., Pandey, R., Leskovec, J. (2015, August). Inferring networks of substitutable and complementary products. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794). ACM.

[4] Grover, A., Leskovec, J. (2016, August). node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864). ACM.