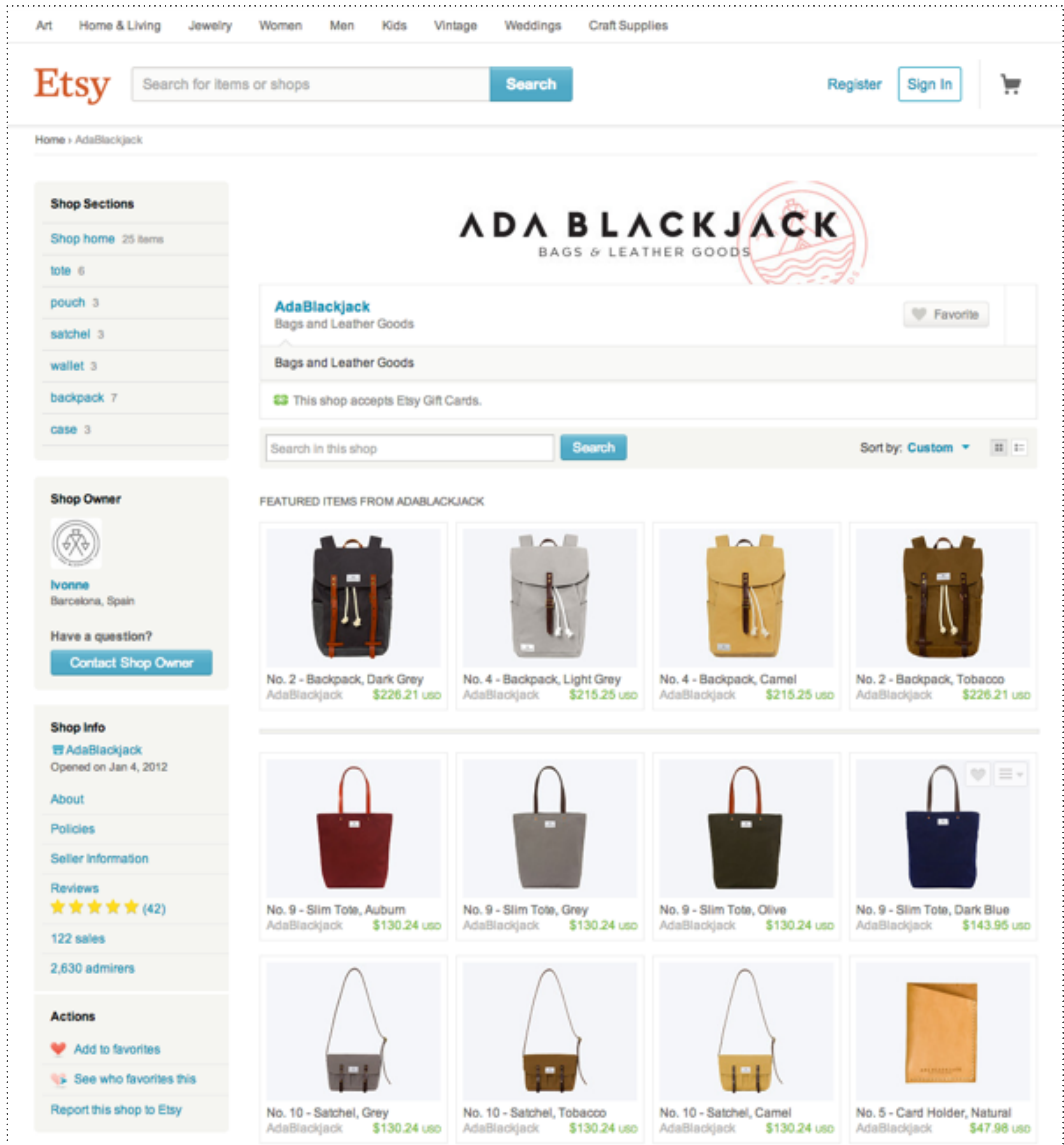
Style in the Long Tail

Discovering Unique Interests with Latent Variable Models in Large Scale Social E-commerce

Diane Hu, **Etsy**
Rob Hall, **Etsy**
Josh Attenberg, **Etsy**

Overview

Etsy is an online marketplace where people connect to buy and sell unique goods: Handmade, vintage, or craft supplies



Ivonne Schippers
Lucas Onofre
Barcelona, Spain

Etsy is an online marketplace where people connect to buy and sell unique goods: Handmade, vintage, or craft supplies

40M	1M	26M	9	200
Users	Shops	Listings	Languages	Countries

How to build recommender systems for such a unique marketplace?

How do people typically use Etsy?

1

Browse (Unintentional)

via the front page, browse pages or the activity feed

Search (Intentional)

for a specific item

2

Favorite

an item or shop, and add to collections with coherent theme/style

Follow

another user with similar style/interest

3

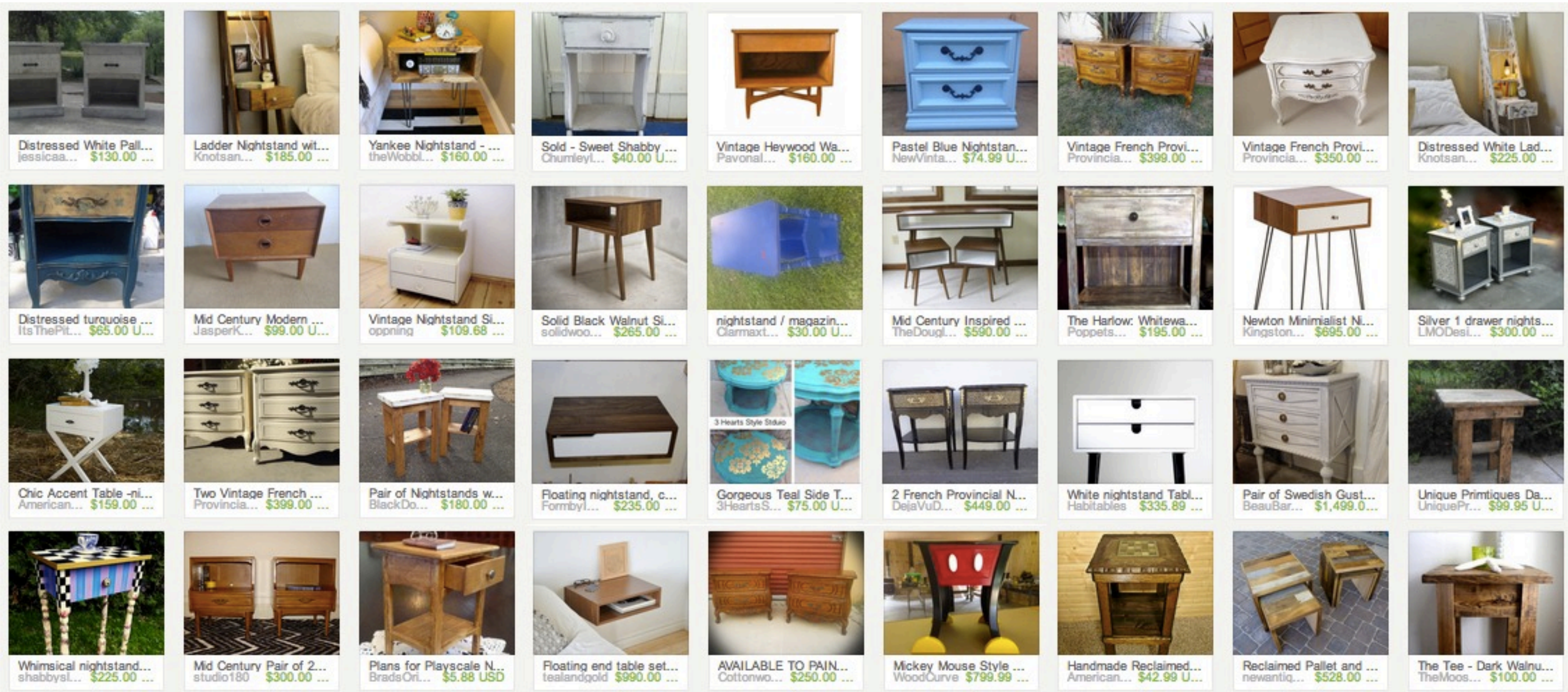
Purchase

an item

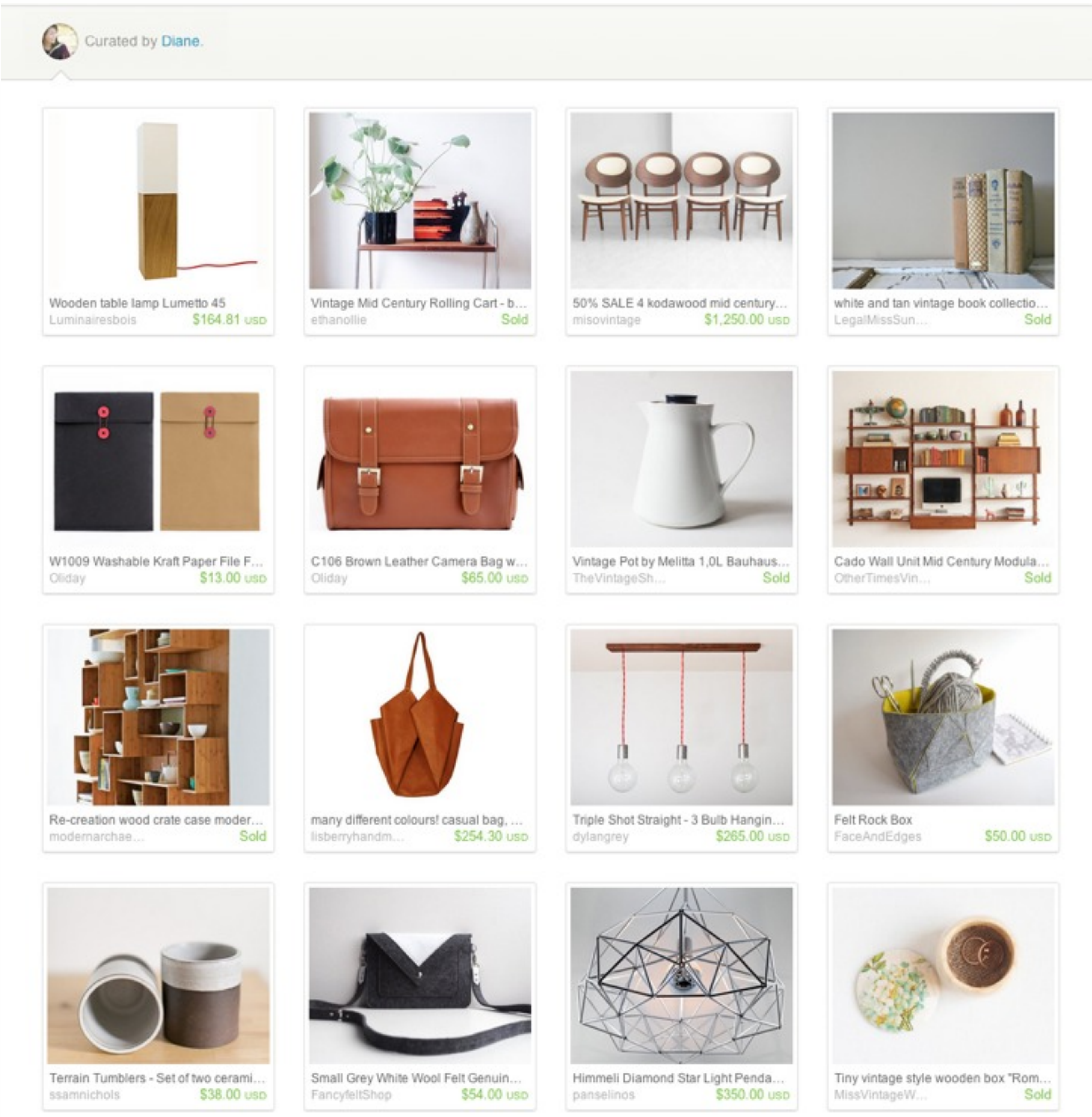
Typical E-Commerce Usage

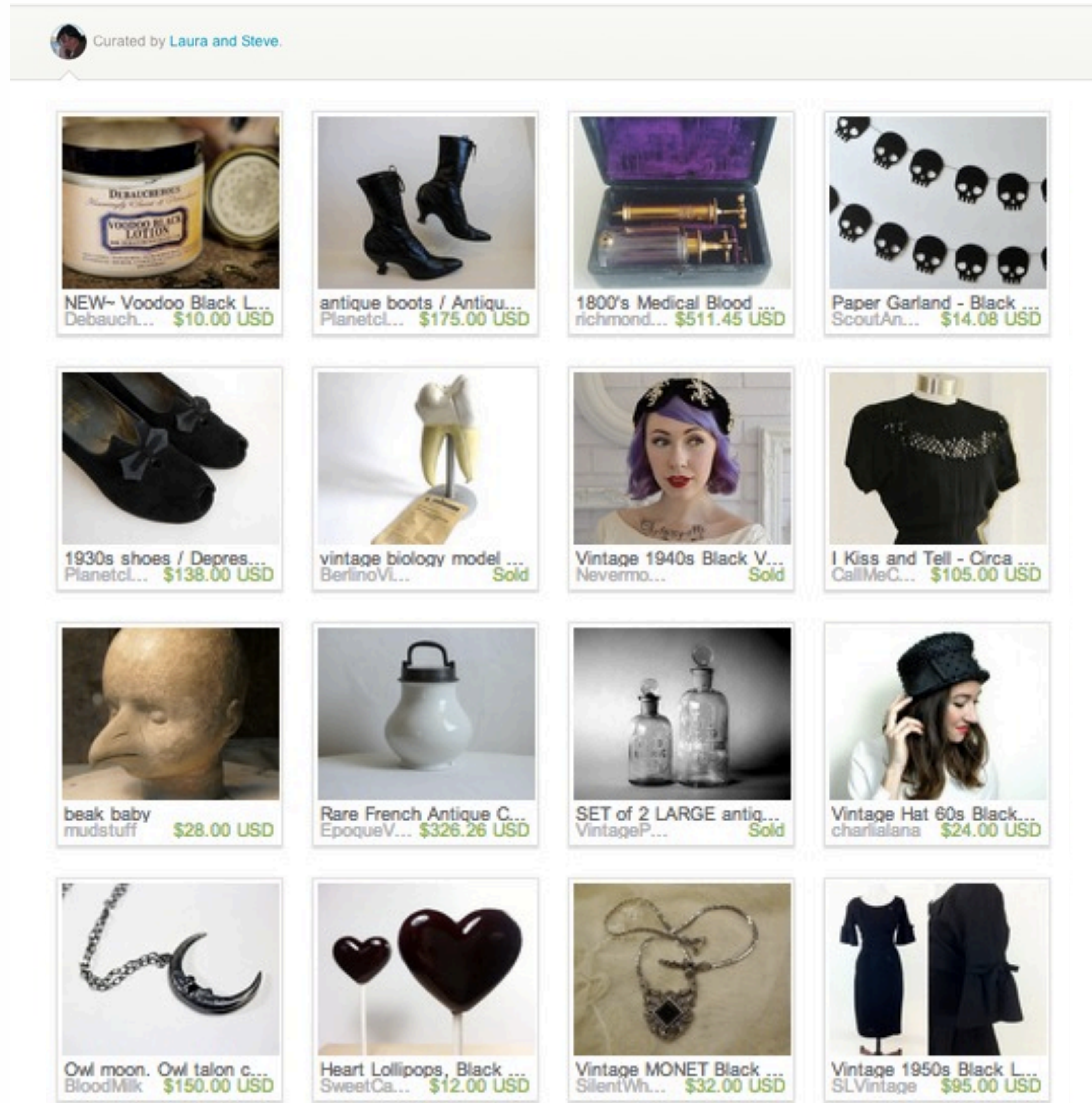
How do people decide what to buy?

Function and style. Example: search results for “nightstand” - 100+ pages



How to describe style?





What Personalization Looks Like on Etsy

Recommendations for multiple intents

Enhance browsing experience, not just purchases

Recommendations for multiple content types

Develop unified method for recommending shops, items, users

Recommendations based on visual styles

Identify user styles and interests in a visually transparent way

Overview

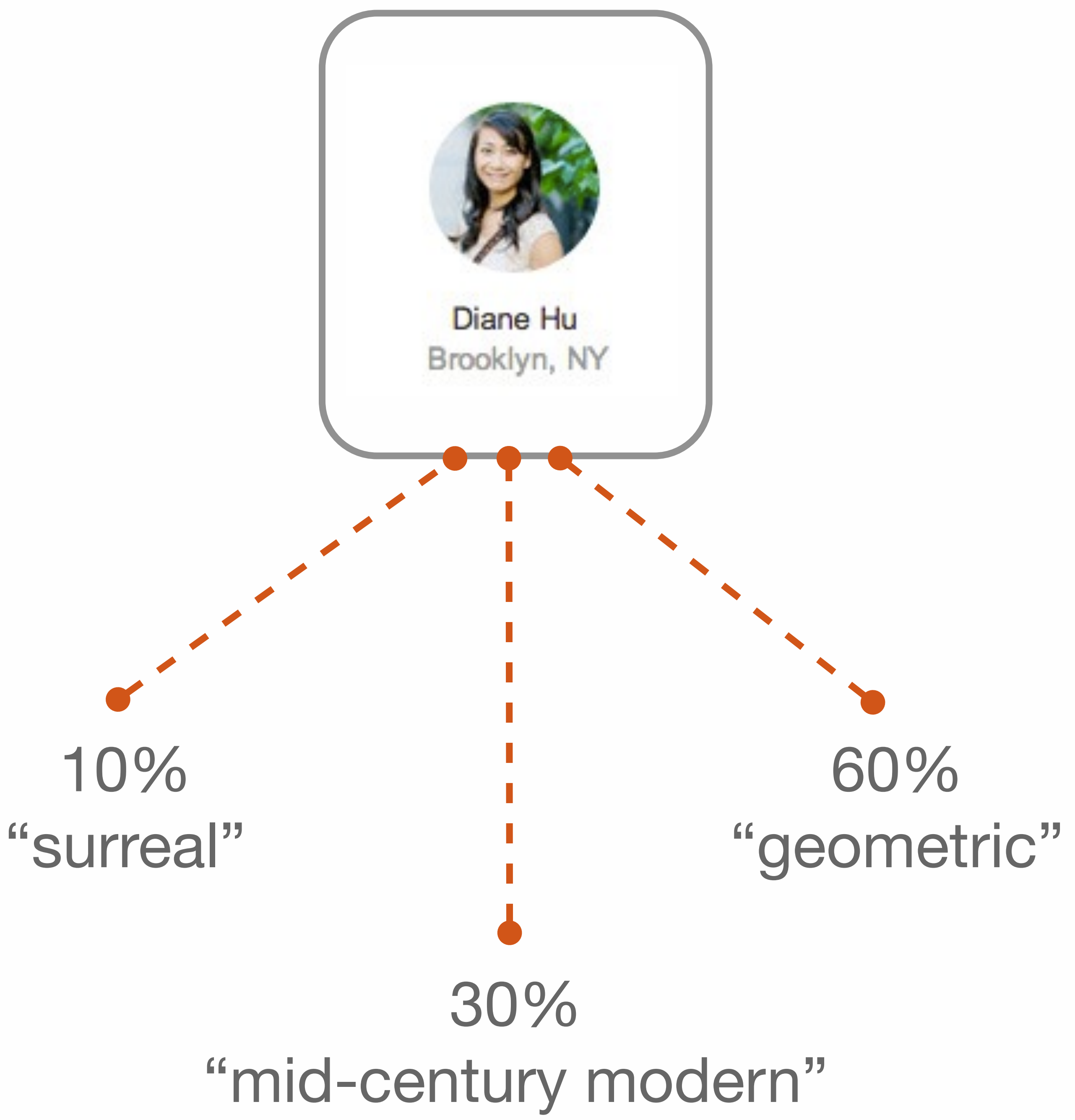
01 | Etsy Overview

02 | Discovering User Styles

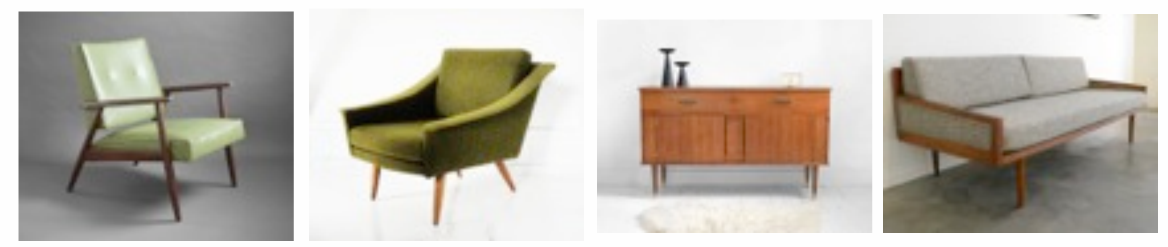
03 | Generating Recommendations

Solution Overview

1 Learn **style profiles** for each user using LDA

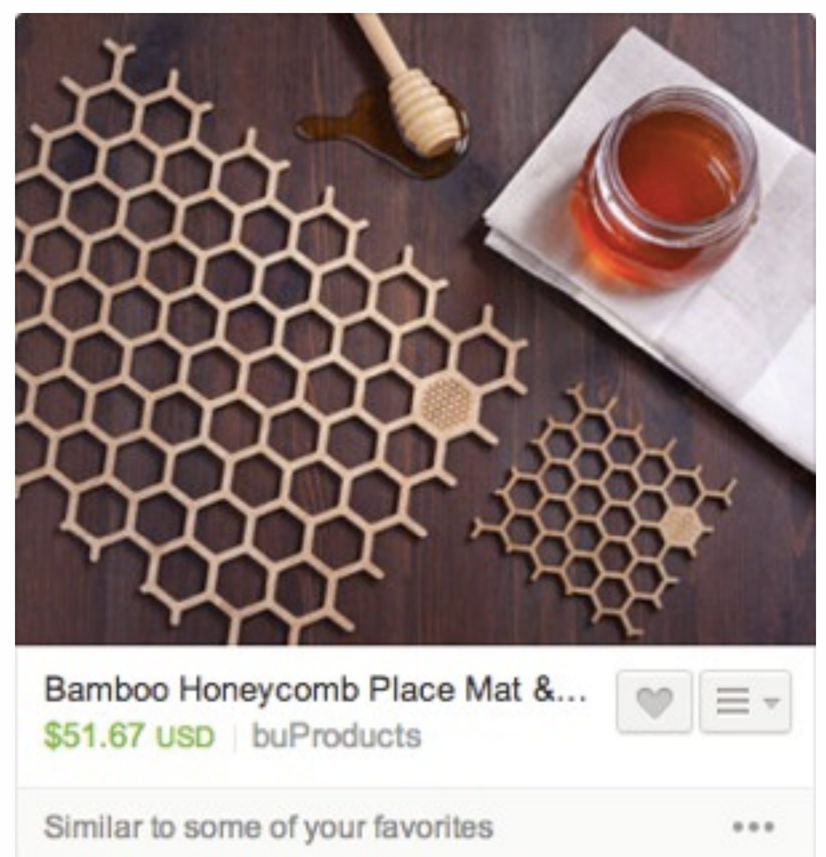


2 Define what each style looks like

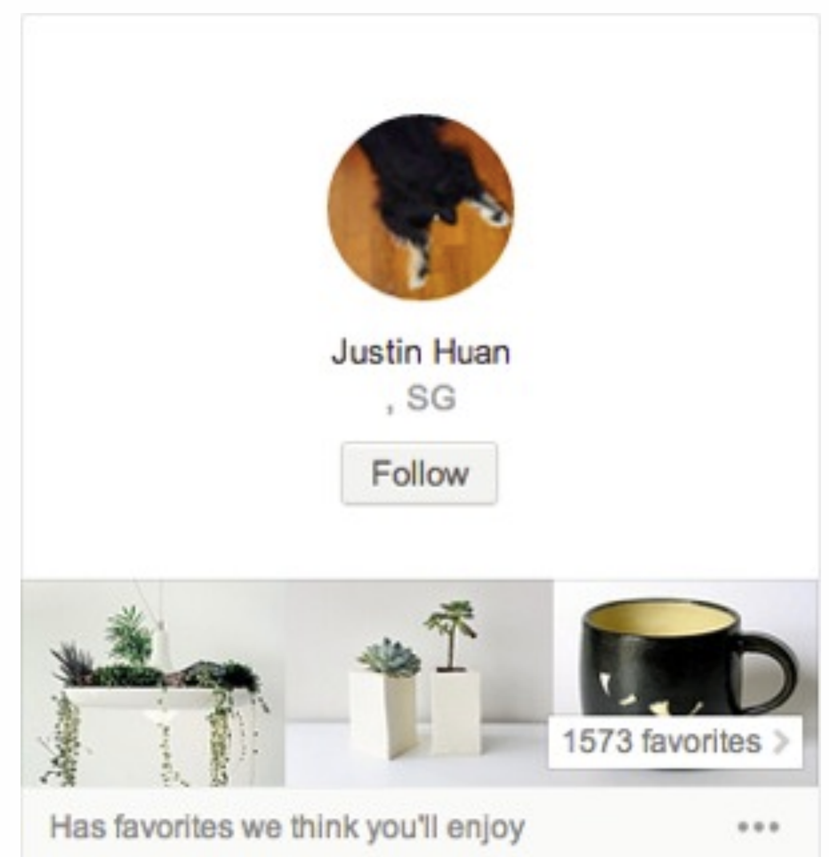


= "mid-century modern"

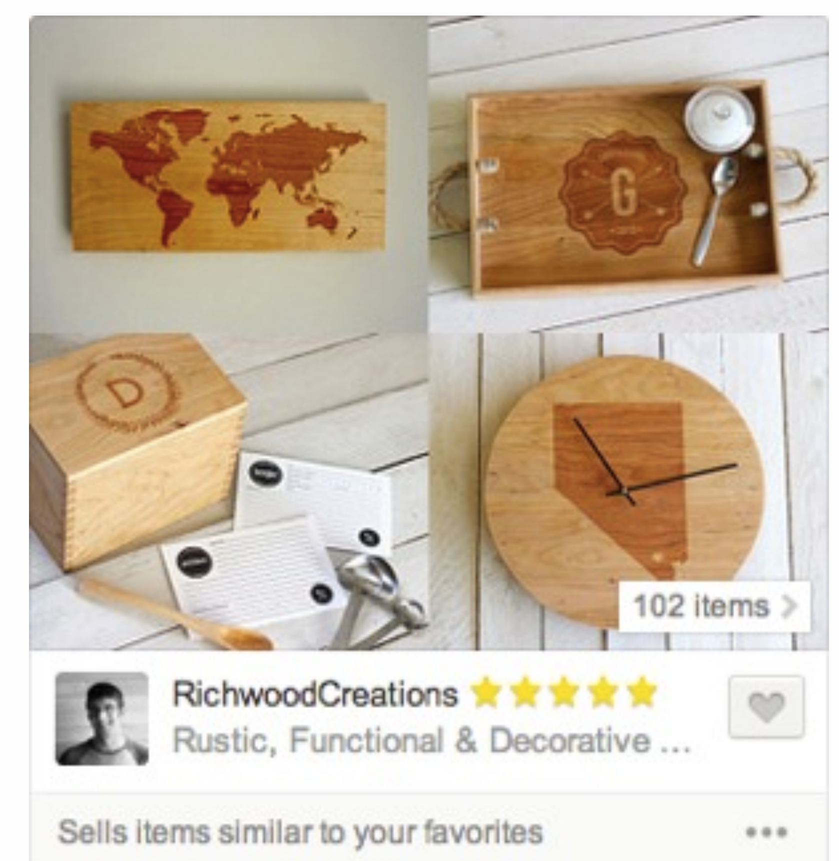
3 User style profiles to generate personalized content



ITEM RECS



USER REC



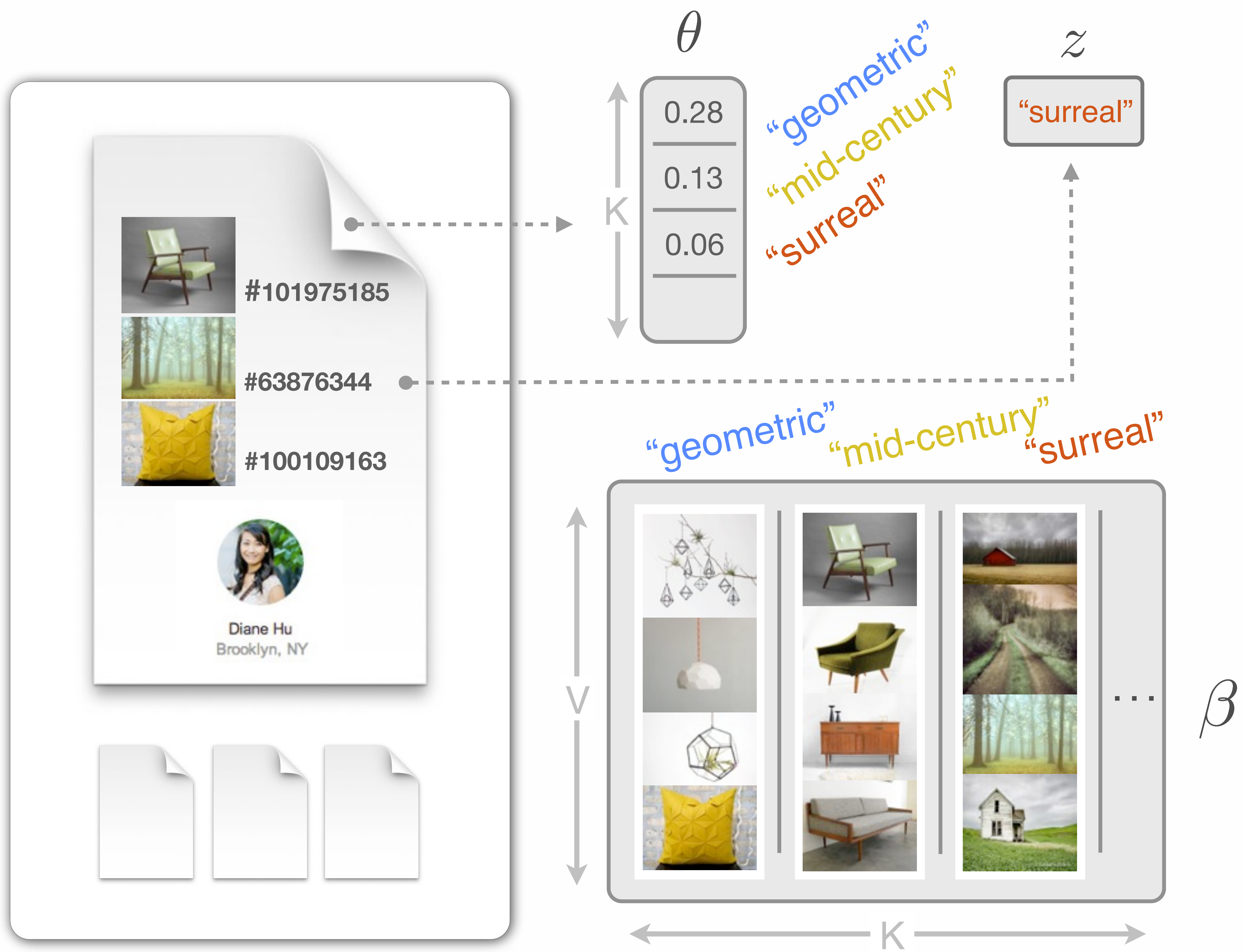
SHOP REC

Latent Dirichlet Allocation (LDA) for Discovering Styles

Assume: Each user's favorited items are generated by this process:

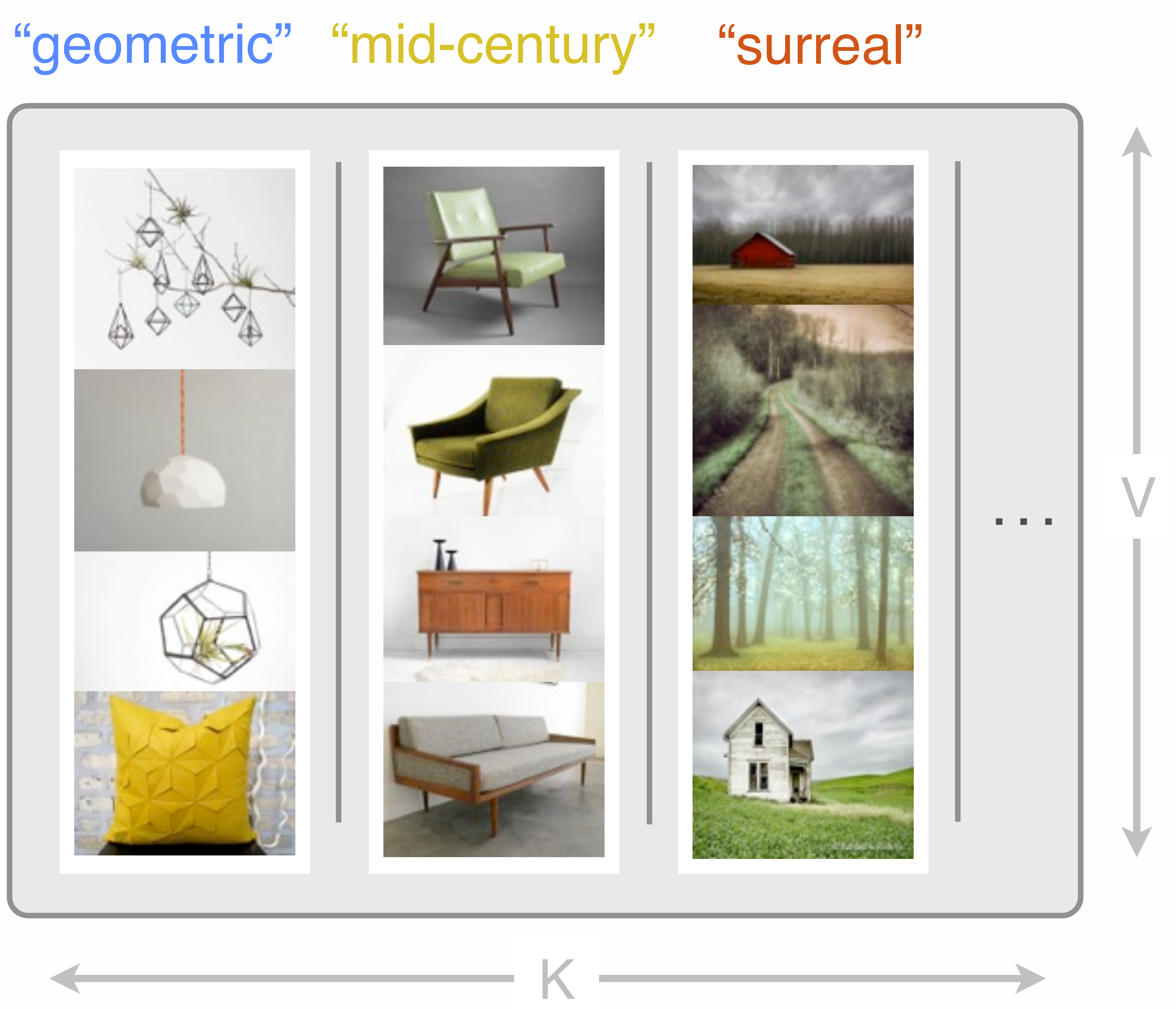
For each user u,

1. Draw a style profile:
   $\theta \sim Dirichlet(\alpha)$
2. For each item, x_n that user u has favorited,
   (a) Draw a style:
   $z_n \sim Multinomial(\theta)$
   (b) Draw an item:
   $x_n \sim Multinomial(\beta_{z_n})$

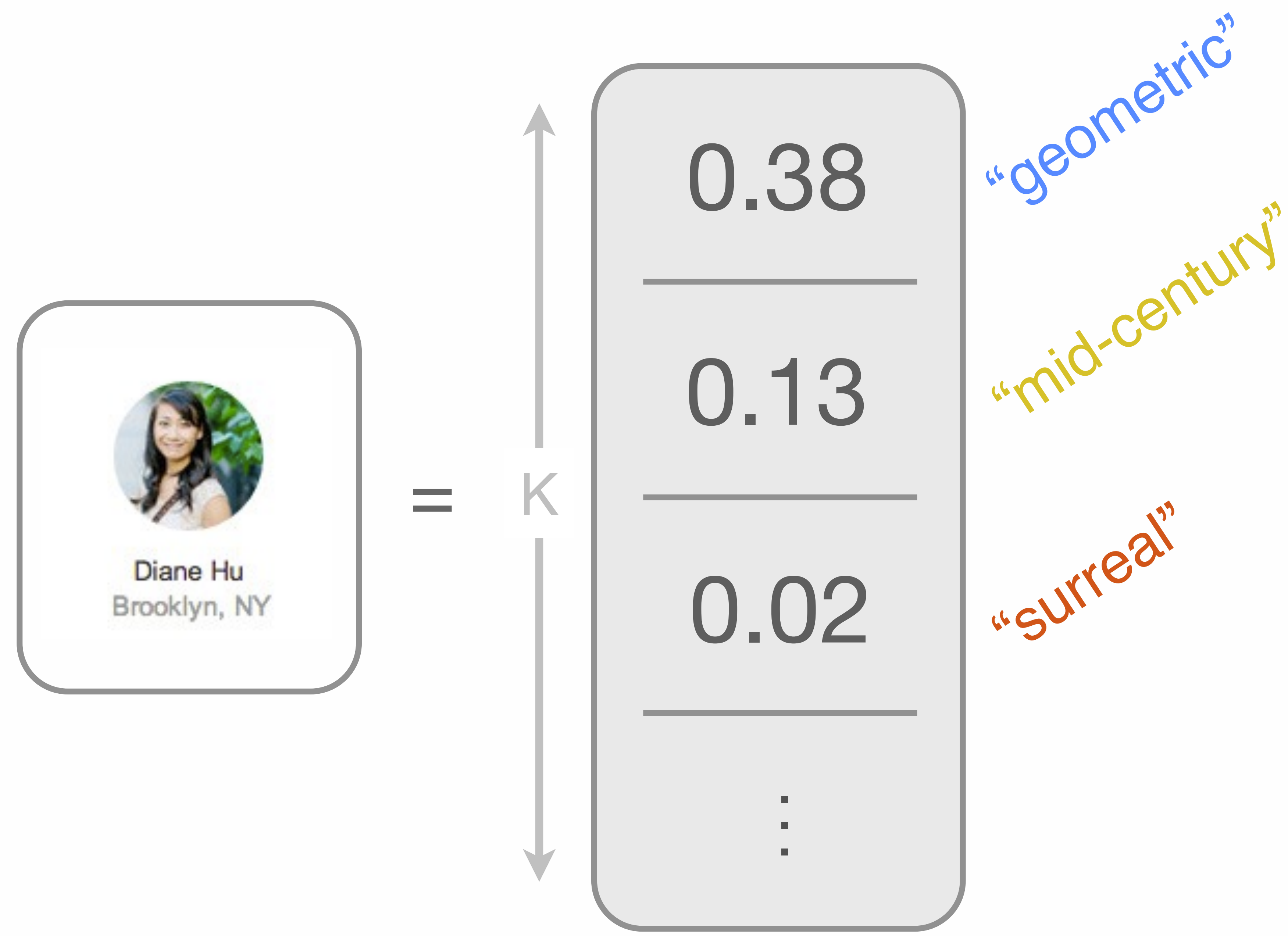


Latent Dirichlet Allocation (LDA) for Discovering Styles

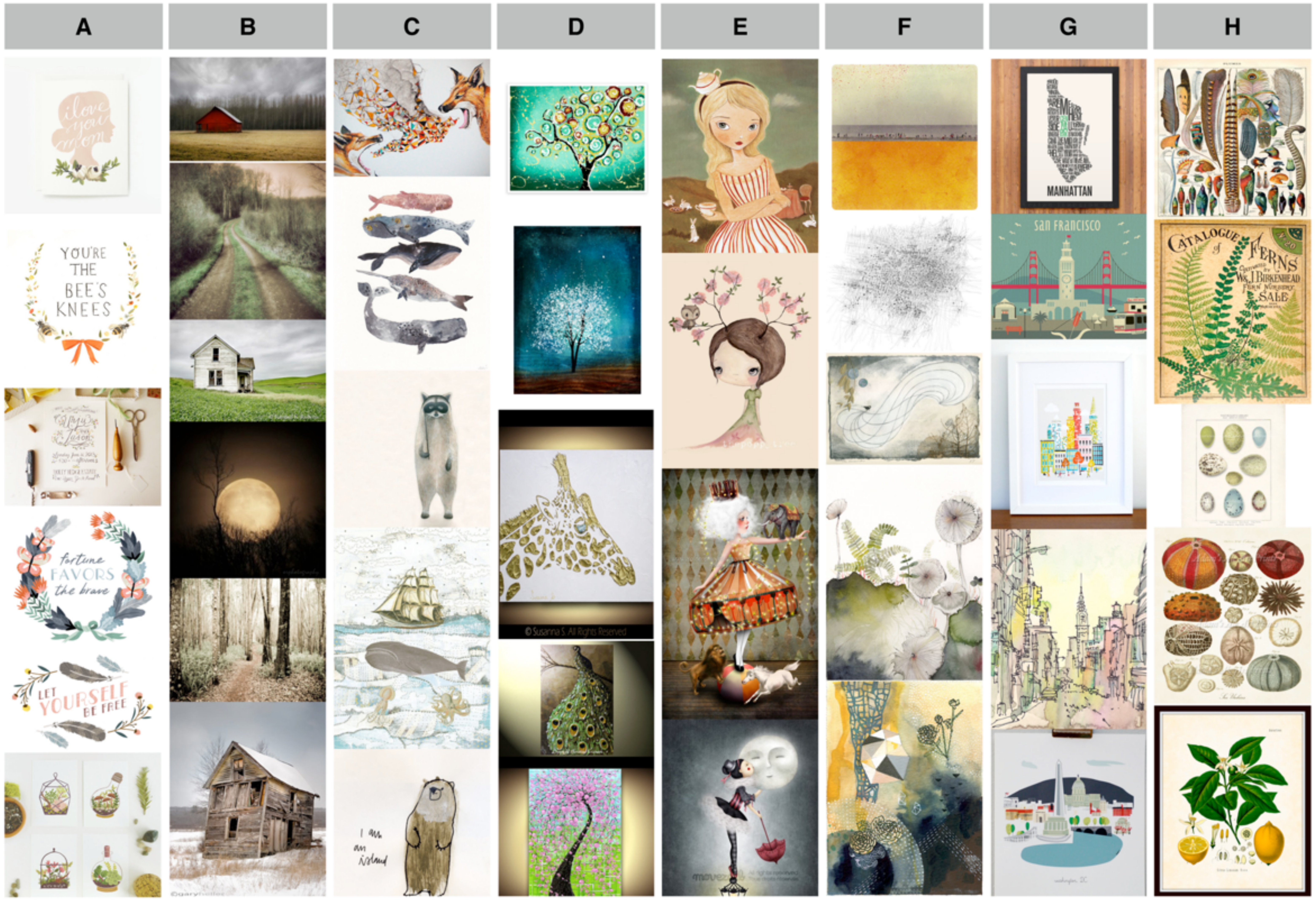
Discover popular styles on Etsy as a distribution over items



Represent each user as a distribution over popular styles, i.e. "**style profile**"



Different styles discovered by LDA



Example of learned styles that contain art prints:

A = Botanical

B = Surreal landscapes

C = Whimsical

D = Acrylic/Abstract

E = French Dolls

F = Whimsical/Abstract

G = Cities

H = Vintage

Different styles discovered by LDA

ANIMALS

TENTACLES

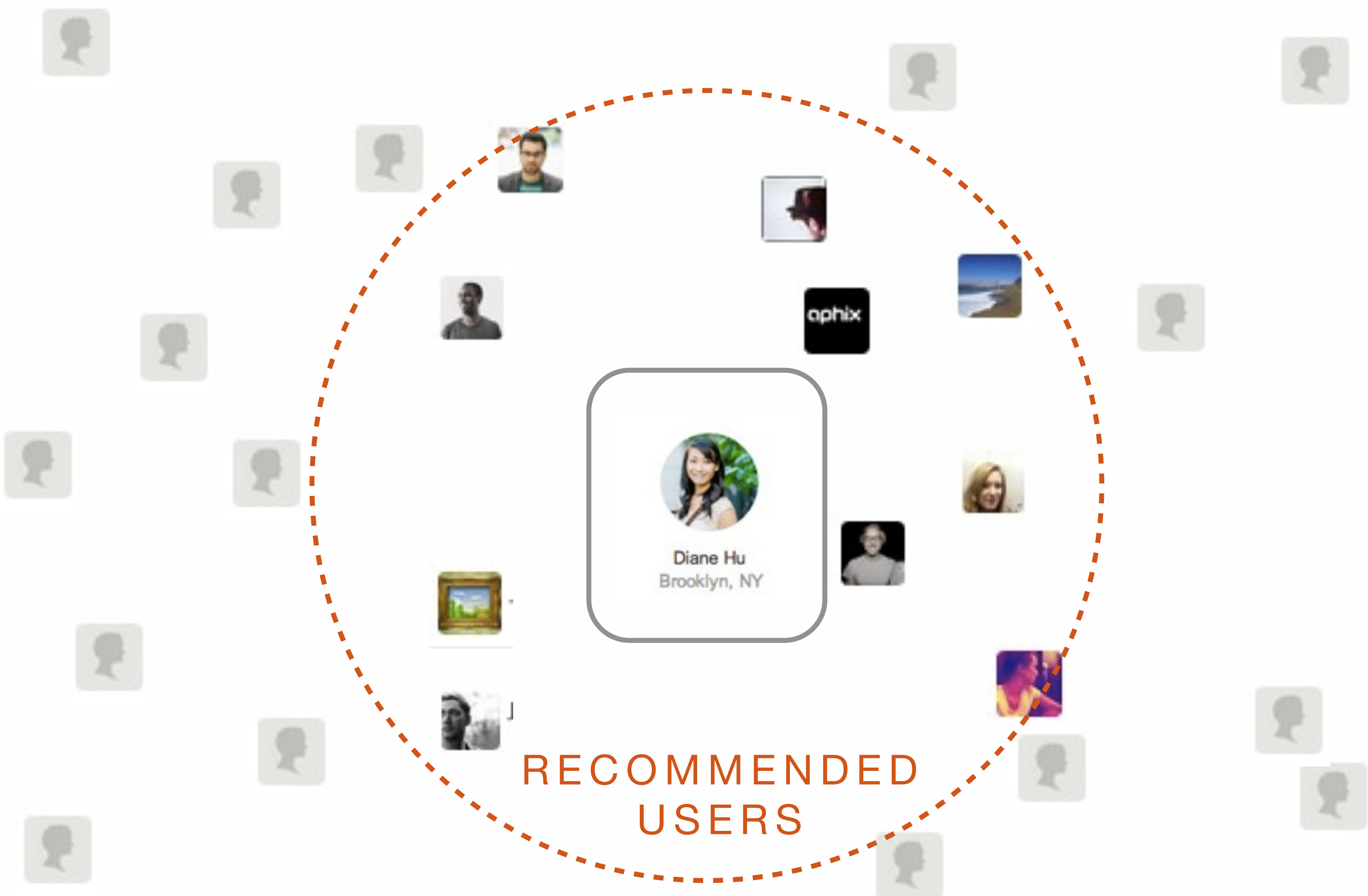
GEOMETRIC

MID-CENTURY MODERN

Overview

User Recommendations

Given that each user has an style profile: Recommend N users with most similar style profiles.



Brute-force K-NN is too expensive. Hash similar users into the same bins, and perform K-NN within each bin

- Locality Sensitive Hashing (LSH). Create hash based on which side of a series of random planes the user falls onto.
- "Top-K" Hashing. Create hash based on set of all pairs of top-k topic indices.

Item Recommendations

Given that each user has an style profile: Sample items most highly weighted styles

USER'S FAVORITES

ITEM RECOMMENDATIONS

STYLE #428

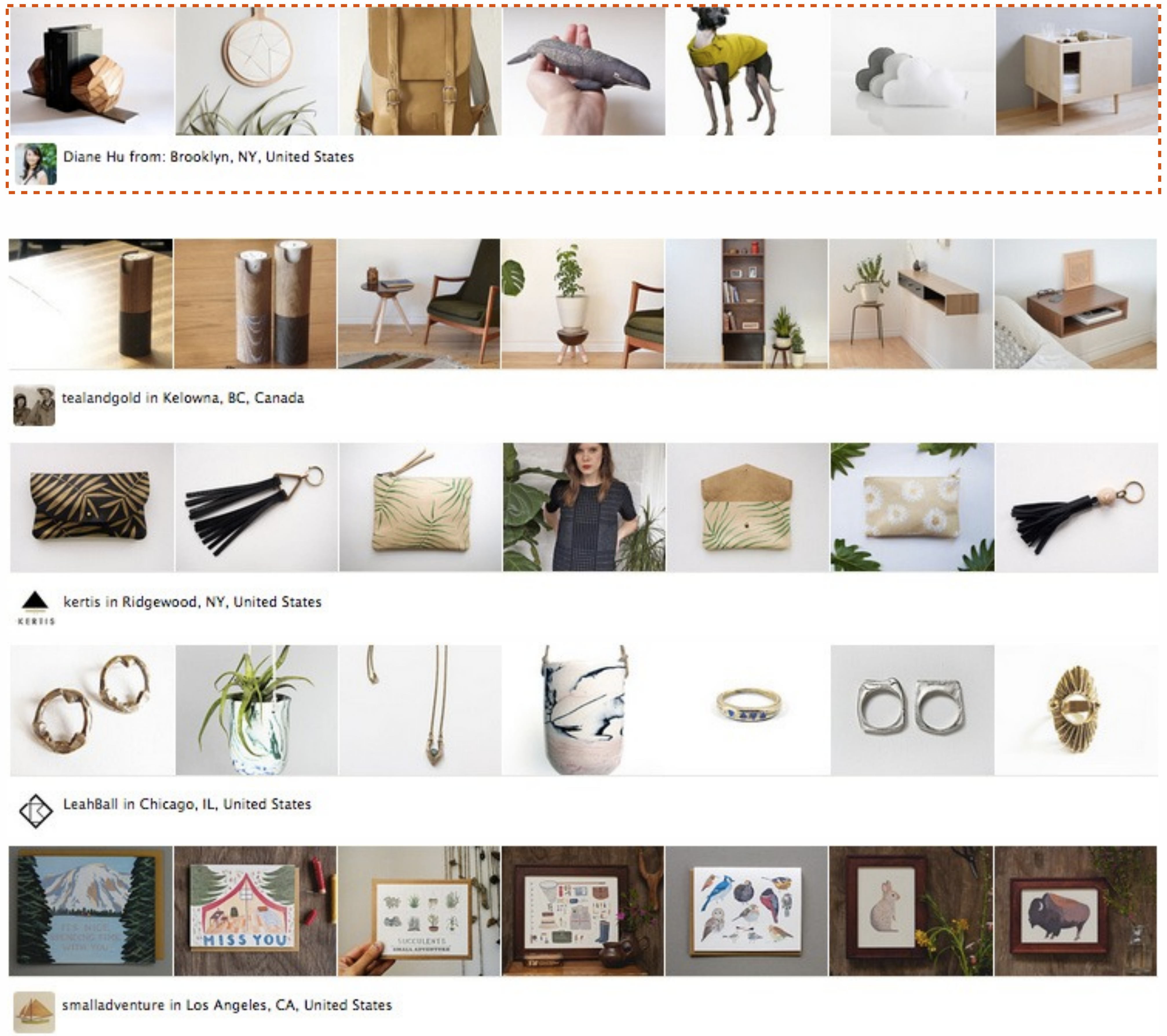
STYLE #54

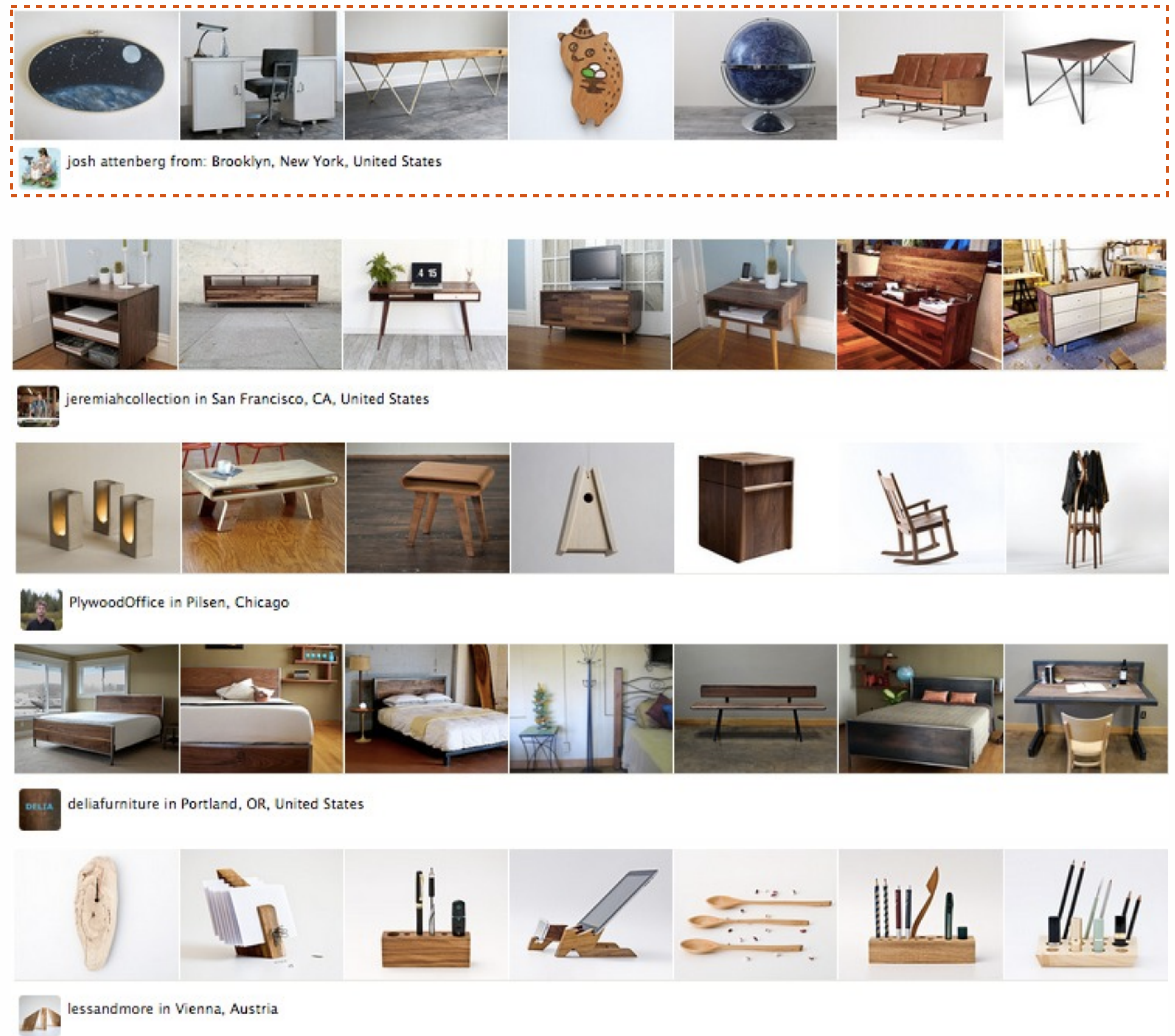
STYLE #655

STYLE #87

Shop Recommendations

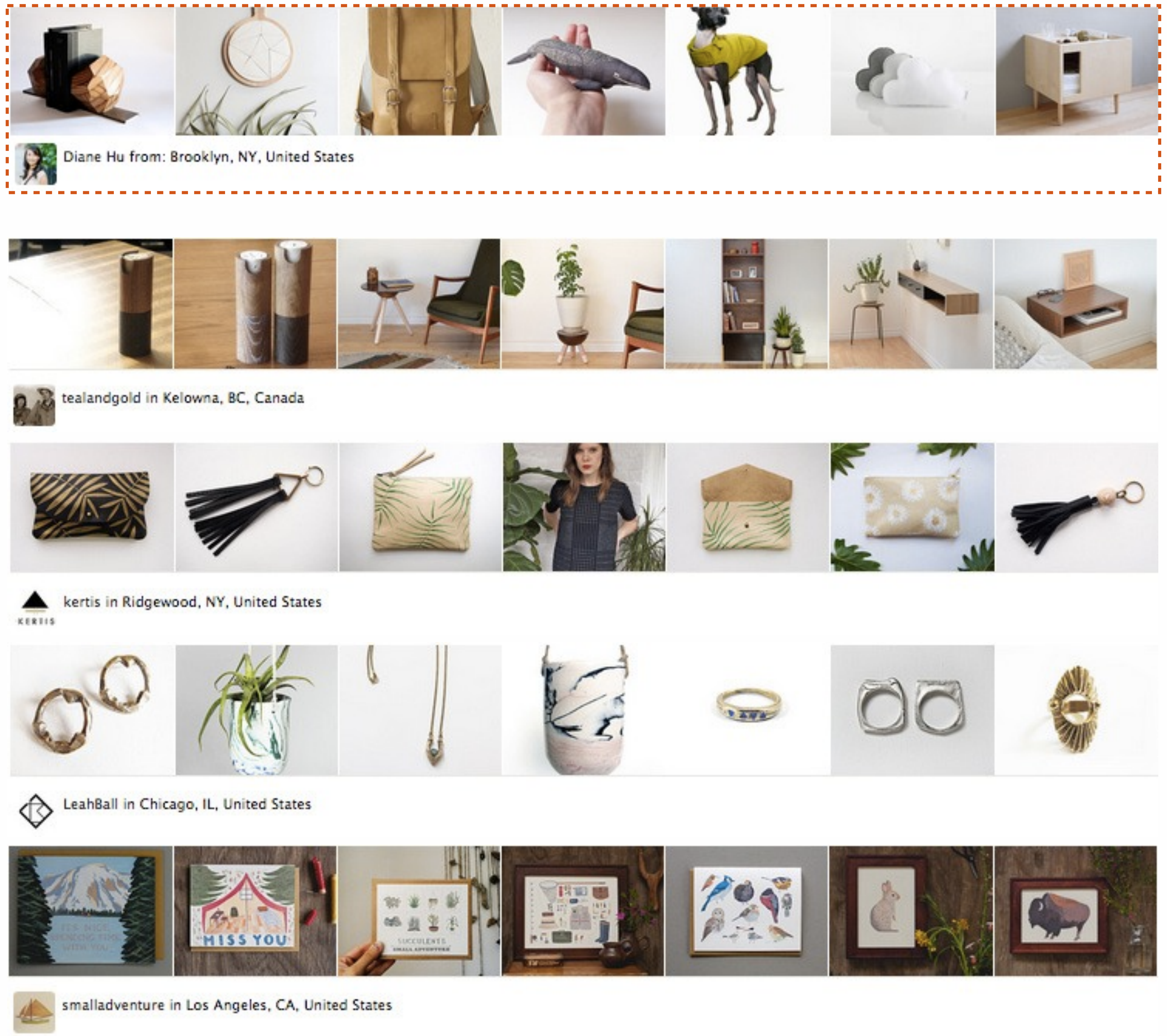
Re-learn topic models substituting item ids with shop ids. Sample shops from highly weighted styles

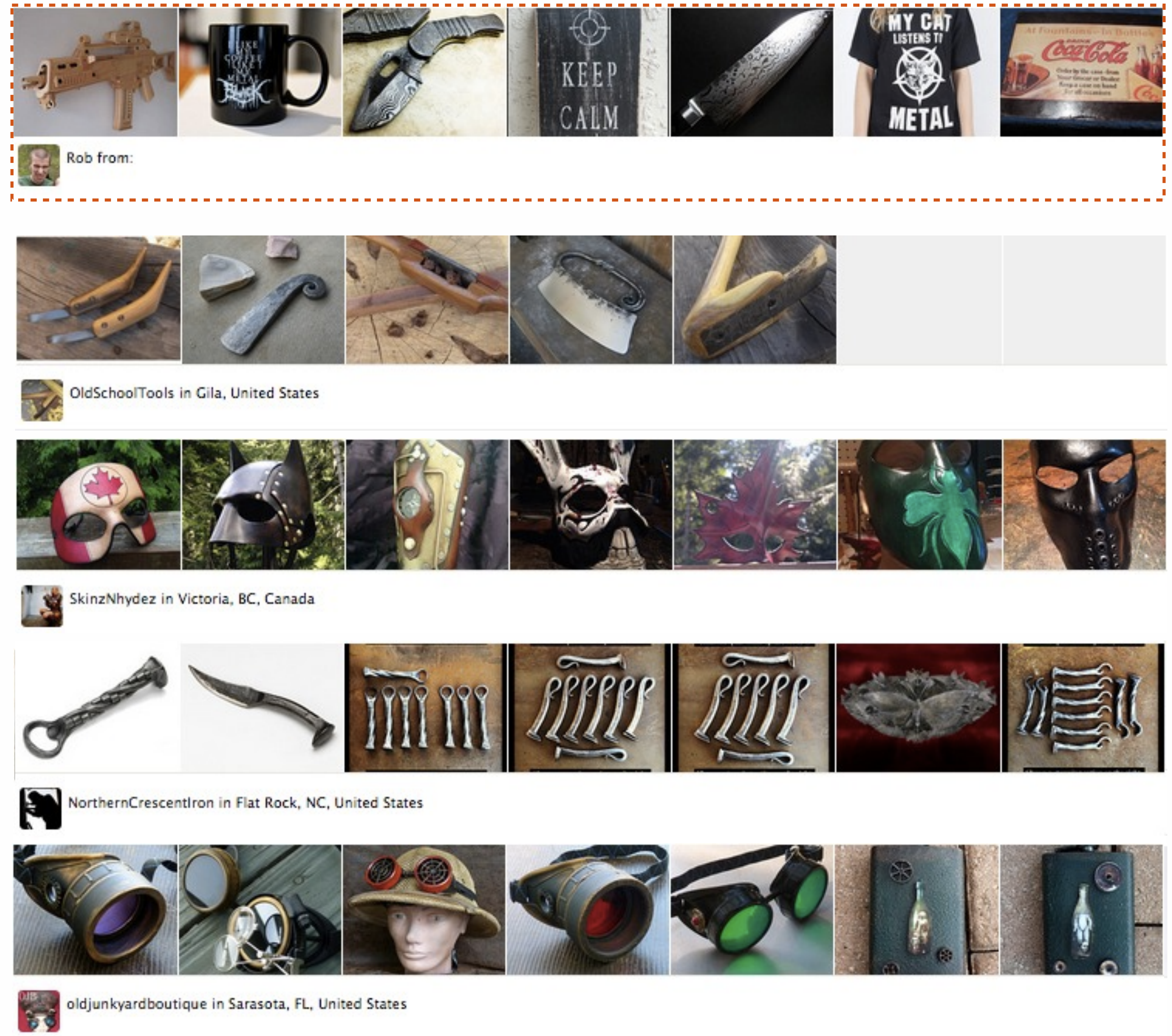




Shop Recommendations

Re-learn topic models substituting item ids with shop ids. Sample shops from highly weighted styles

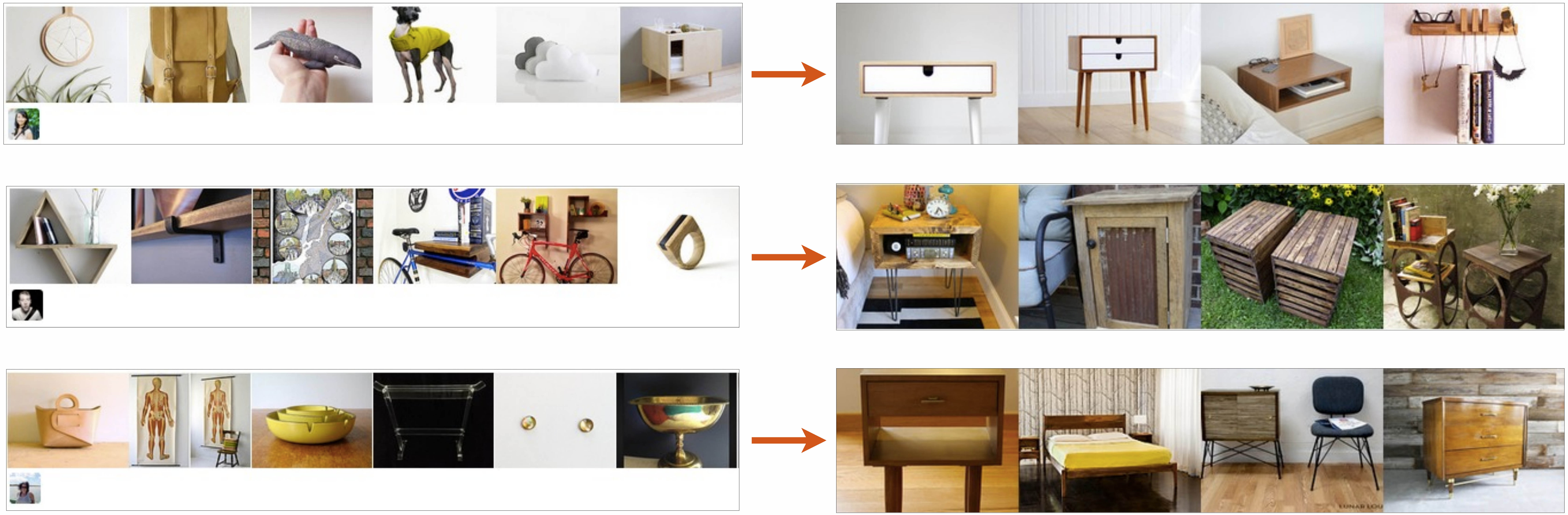




Recommending Styles within Categories

Find how overall styles translate into specific categories

USER'S RECENT FAVORITED ITEMS | **RECOMMENDATIONS IN "NIGHTSTANDS"**

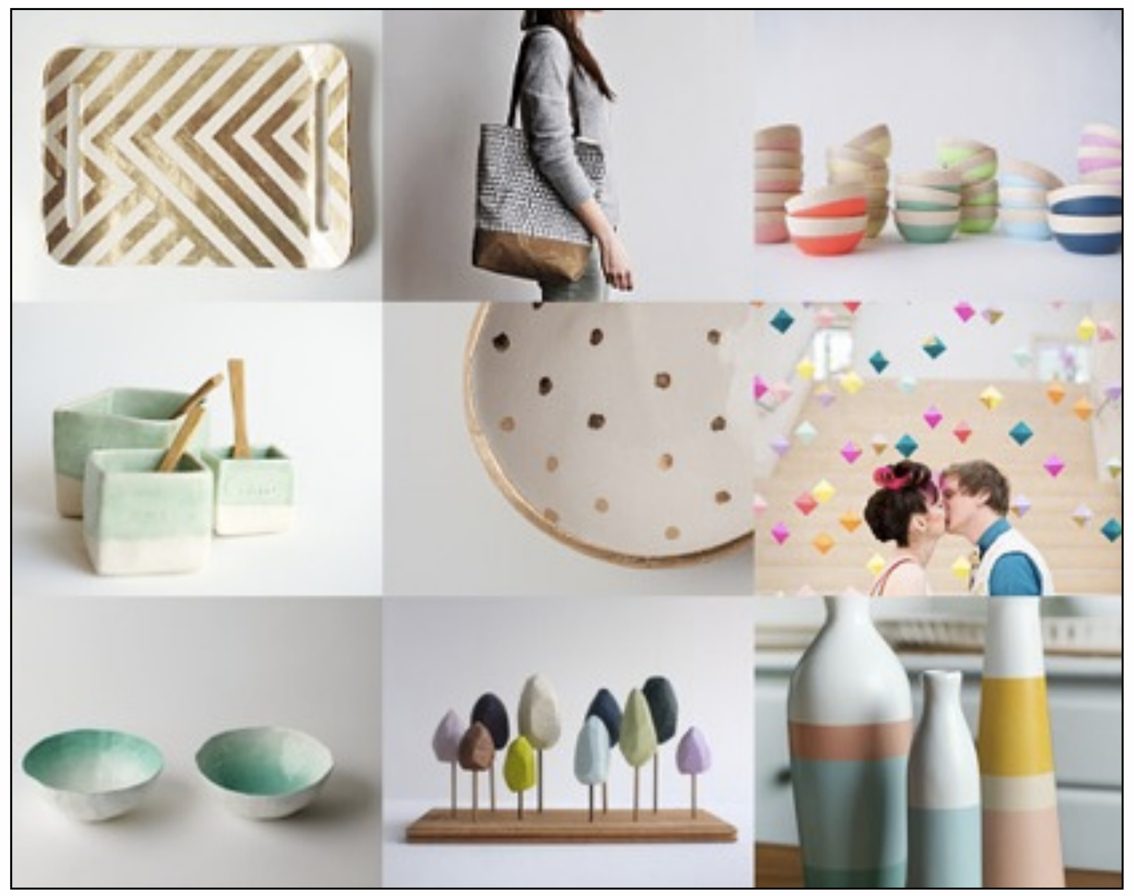


Visualizing Related Topics

Learn topic correlations from users' style-profiles.

STYLE #1

STYLE #2



STYLE #3

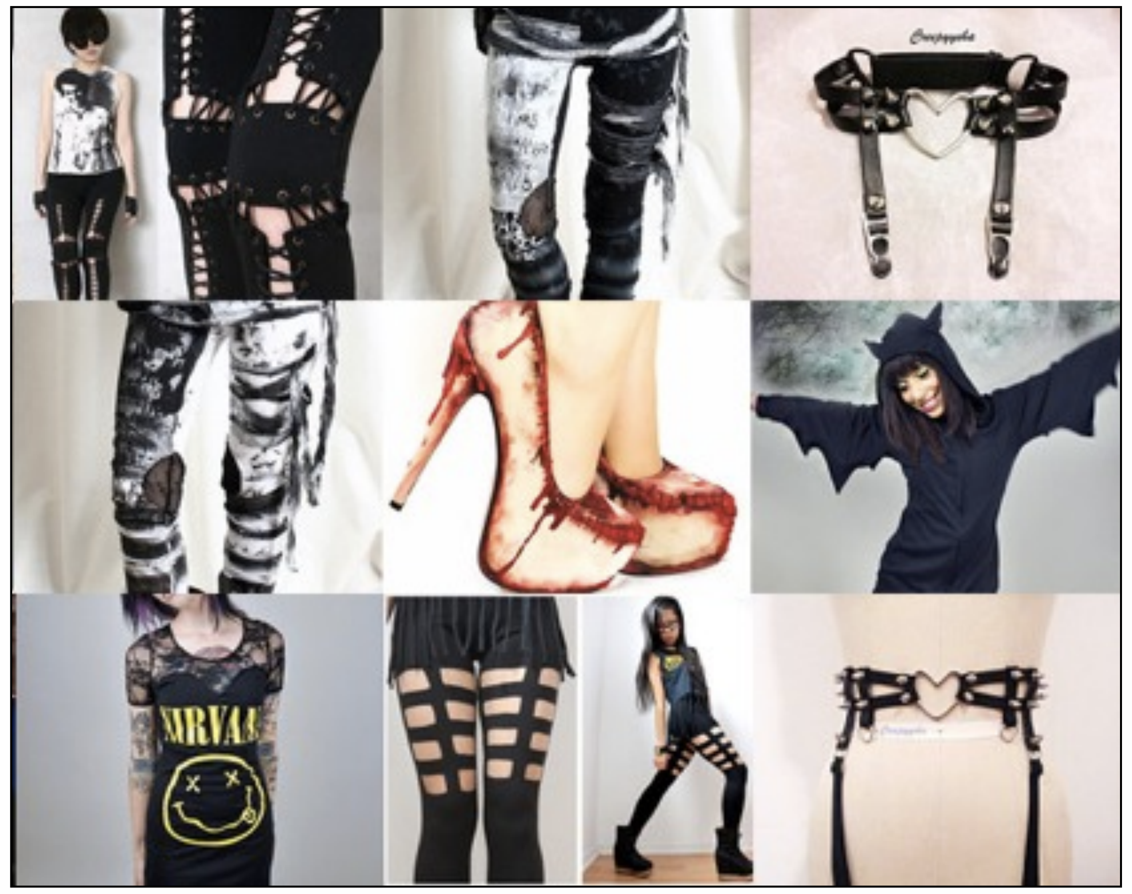
STYLE #4

"Bright, Whimsical"

Visualizing Related Topics

Learn topic correlations from users' style-profiles.

STYLE #1



STYLE #2

STYLE #3

STYLE #4

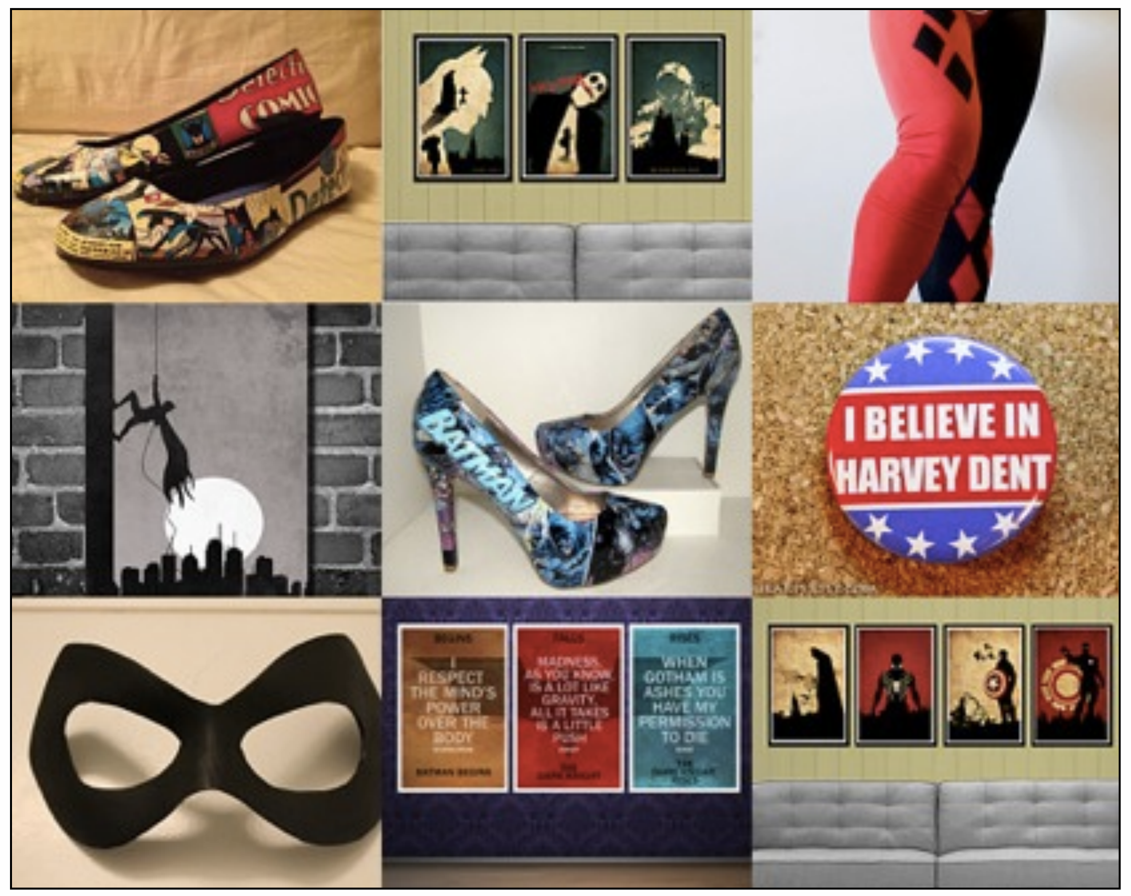
"Gothic Punk"

Visualizing Related Topics

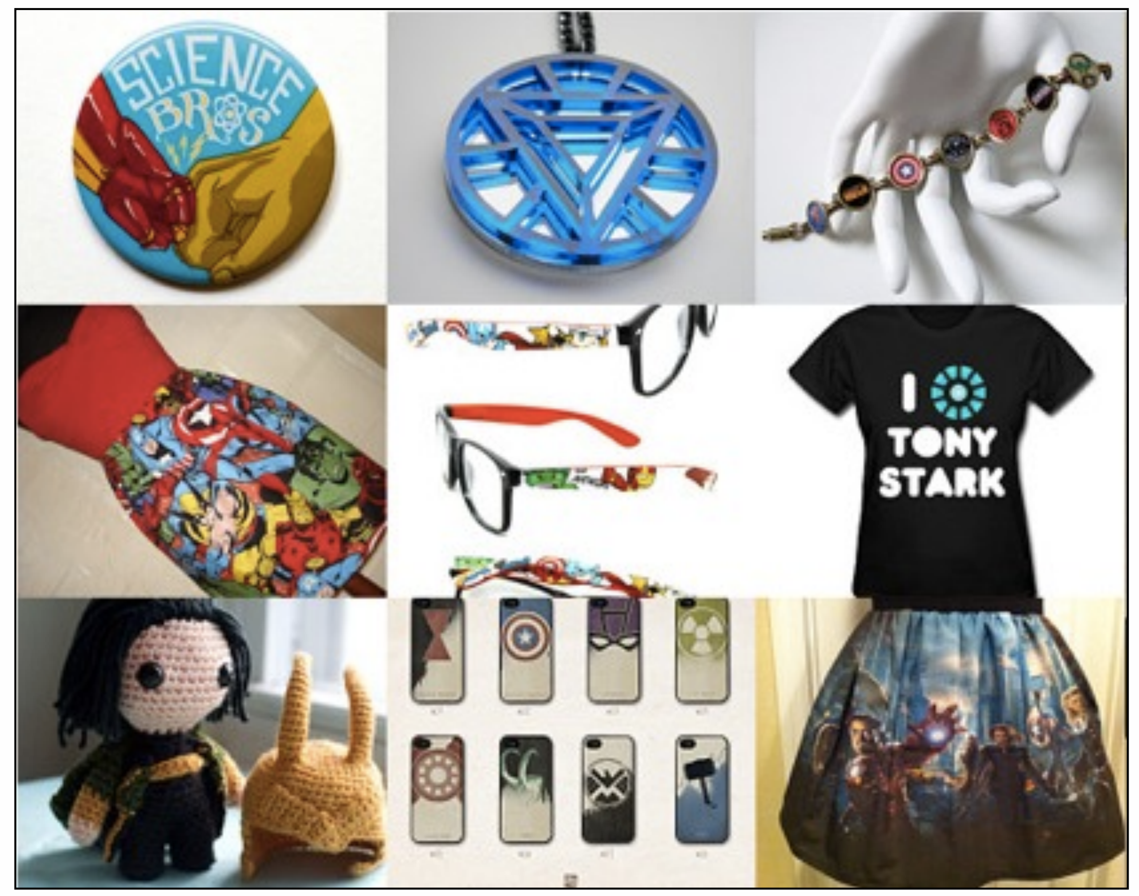
Learn topic correlations from users' style-profiles.

STYLE #1

STYLE #2



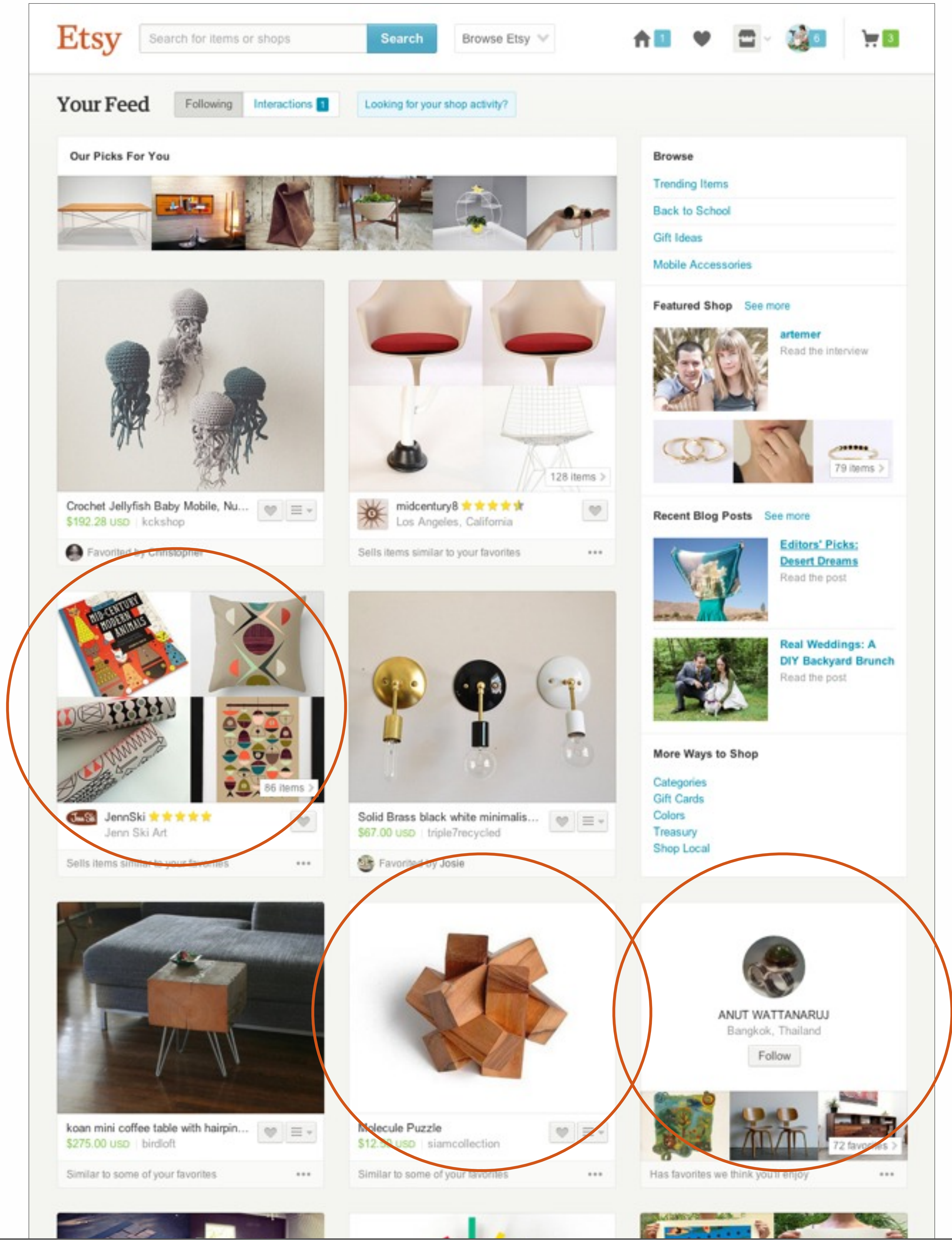
STYLE #3



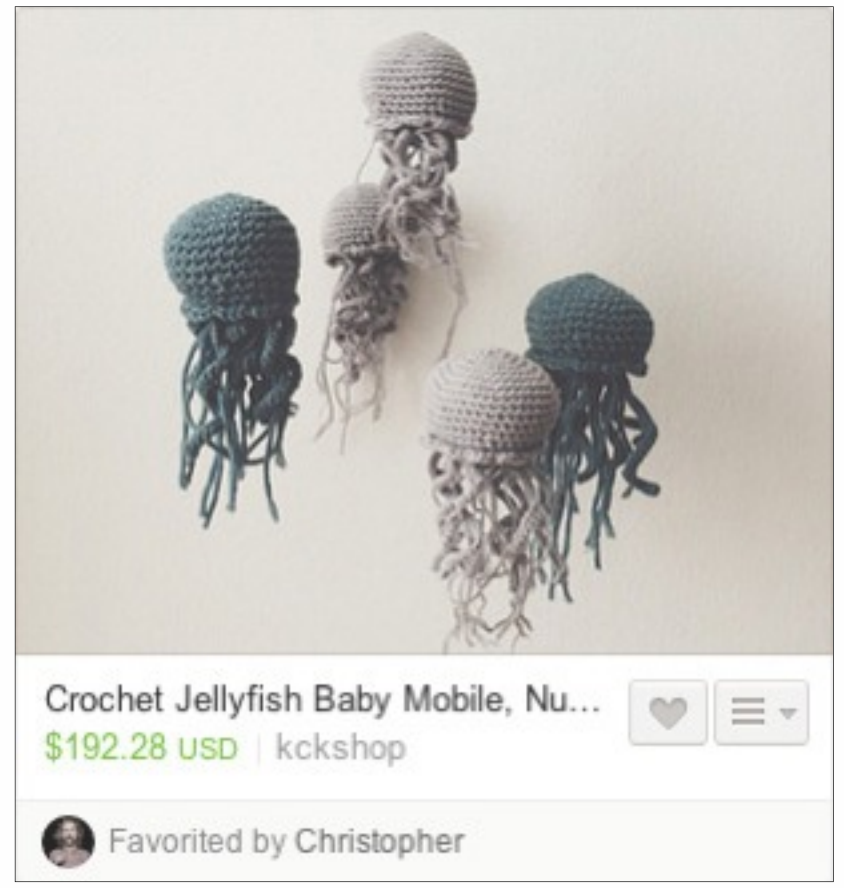
STYLE #4

"Sci-fi/Fantasy"

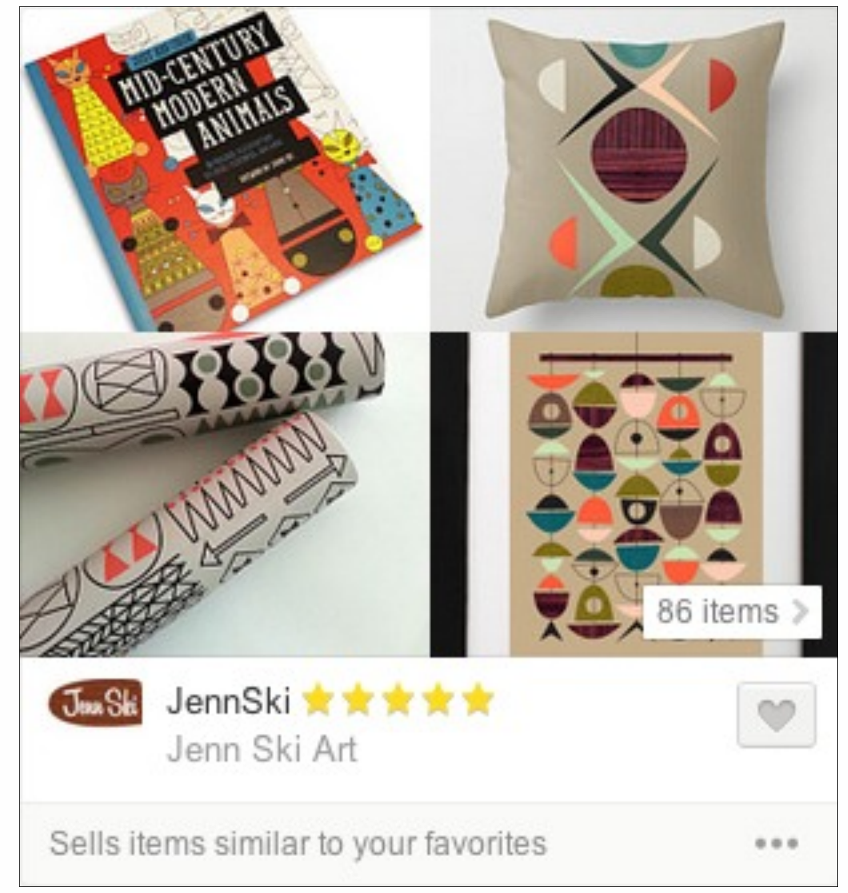
Recommendations in the Activity Feed



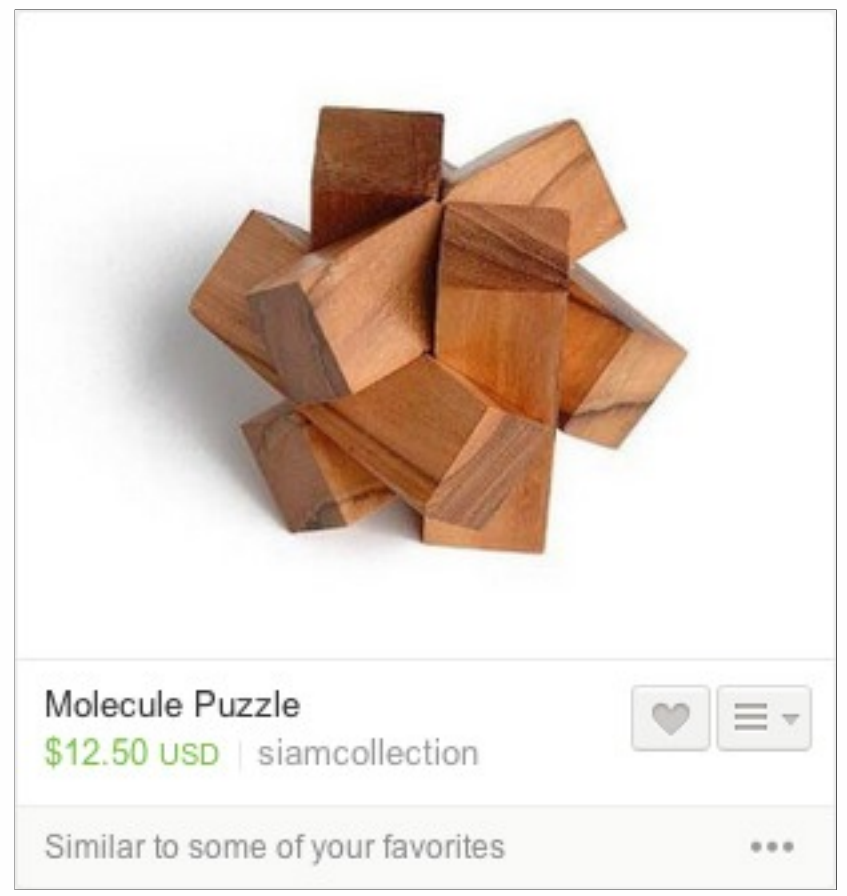
ORGANIC



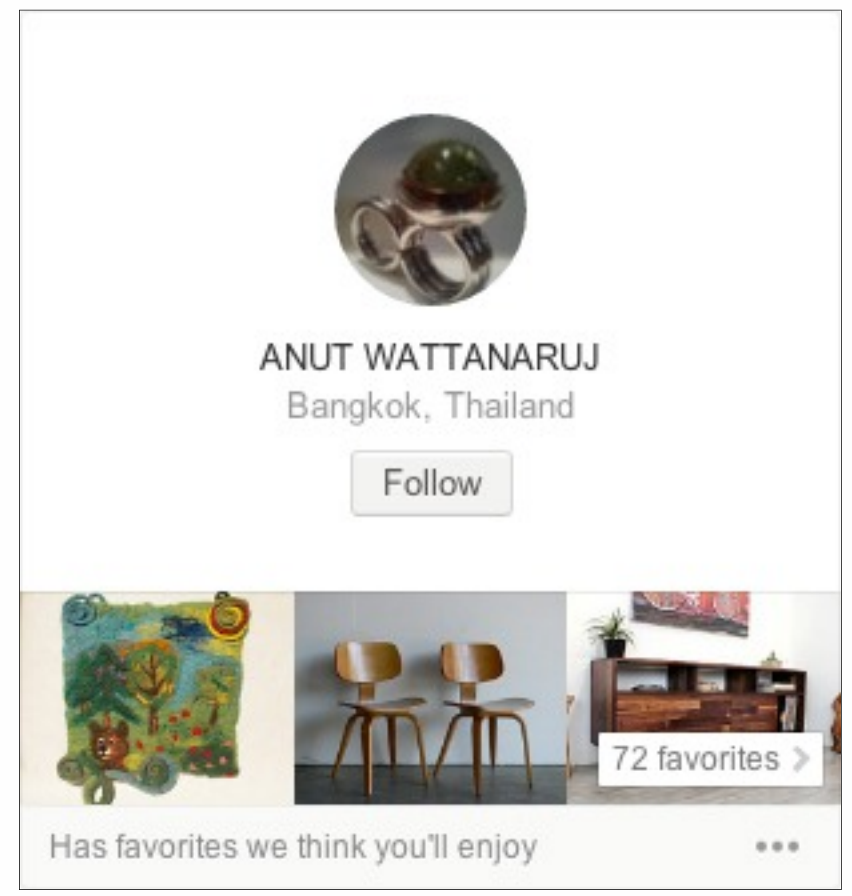
SHOP REC



ITEM REC



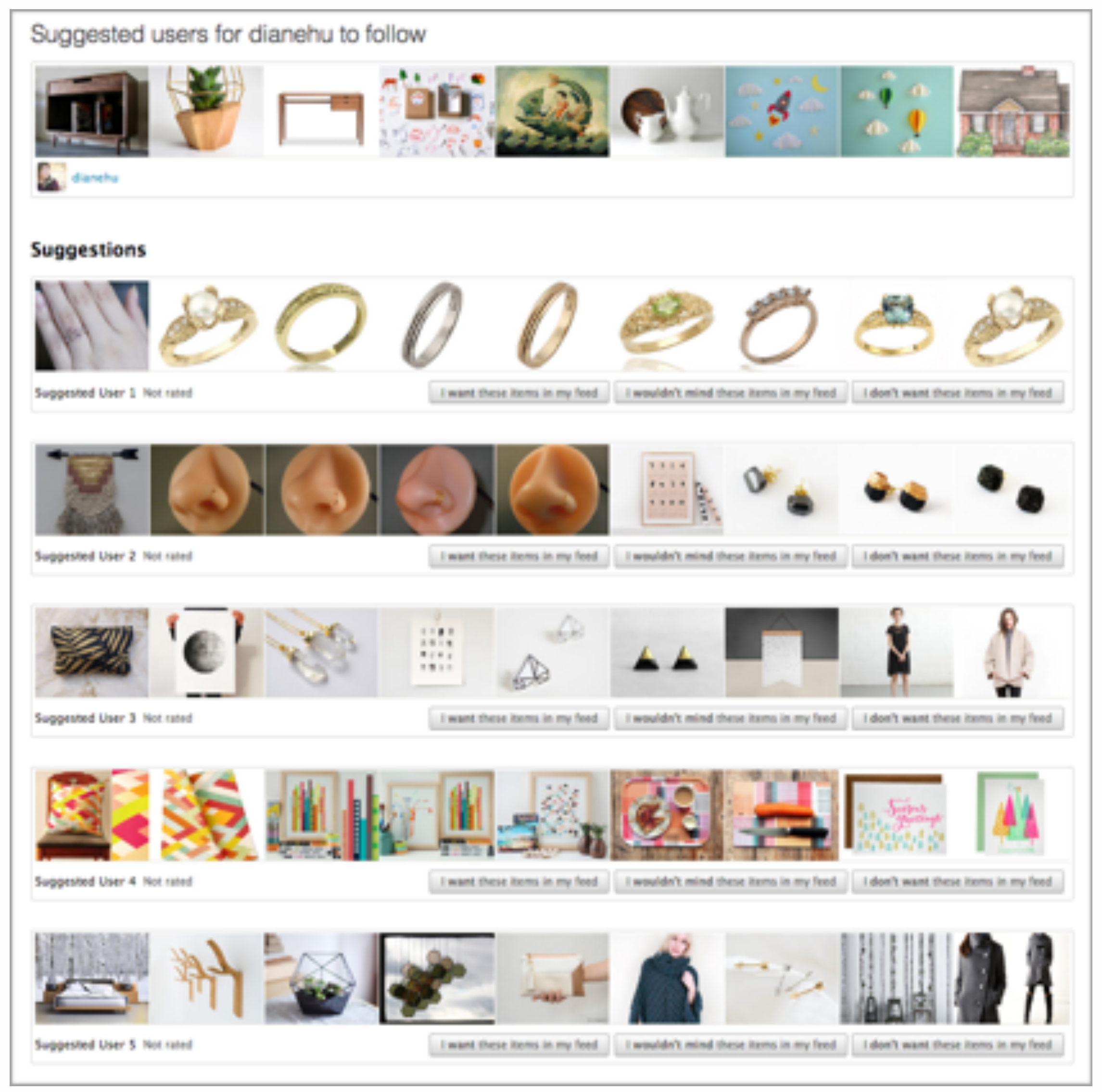
USER REC



User Recommendation Experiments

Side-by-Side User Study

- Randomly interleave user recs from 3 algorithms: (1) LDA, (2) TF-IDF w/ Cosine Similarity, (3) Triadic Closure
- User rated each recommendation positive, neutral, negative
- LDA was overwhelming winner



A/B Testing in Activity Feed

Phase One:

- LDA vs. No recs
- Significantly increased all business metrics

Phase Two:

- Different variants of LDA vs. Matrix Factorization (using Stochastic SVD)
- Matrix factorization and LDA comparable across business metrics

Conclusion

What We Did

- Identify styles across Etsy as a visual experience
- Generate style profiles that are visually transparent and capture diverse taste
- Build large-scale recommender systems:
  - for multiple content types
  - for enhancing browse experience
- Improve key business metrics

More Details On

- System/hard-ware set-up
- Scaling algorithms to ~40M users
- Experimental set-up and outcomes
- Product design for recommendations

Etsy