ROBERT H. SMITH SCHOOL OF BUSINESS

IDENTIFYING MARKET STRUCTURE: A DEEP NETWORK REPRESENTATION LEARNING OF SOCIAL ENGAGEMENT

Kunpeng (KZ) Zhang

kpzhang@umd.edu

kpzhang.github.io

Joint work with **Yi Yang** (HKUST) & **P. K. Kannan** (University of Maryland)

What is competitive market structure?

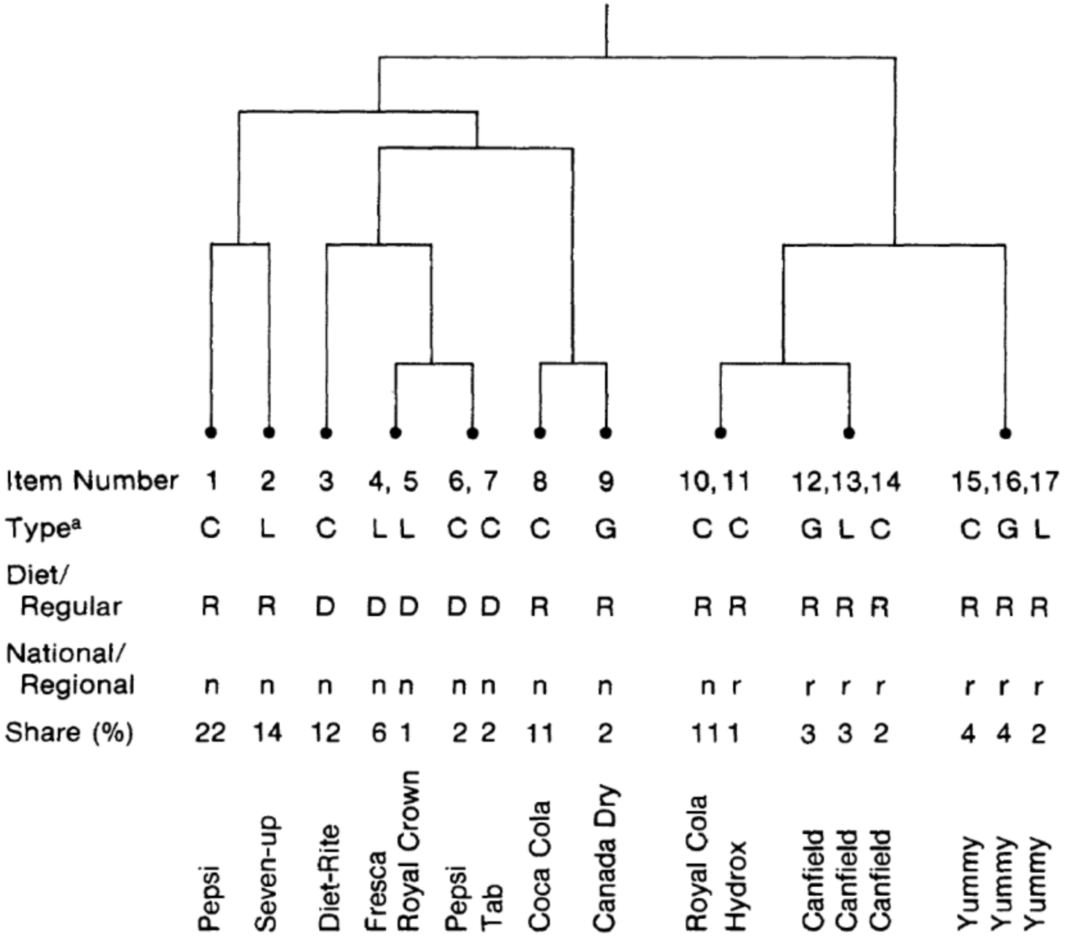
- Understanding the extent of competition among brands in a product-market
- Identifying sub-markets with the market, where competition within a sub-market is much stronger than competition across sub-markets
- Given a focal brand, identifying brands in the market that compete very closely with it as compared to other brands

Early market structure research

- Rao and Sabavala (1981)
- Input: panel data of consumer purchases/switching
- Similarity data using brand switching matrix
- Hierarchical clustering

FIGURE D

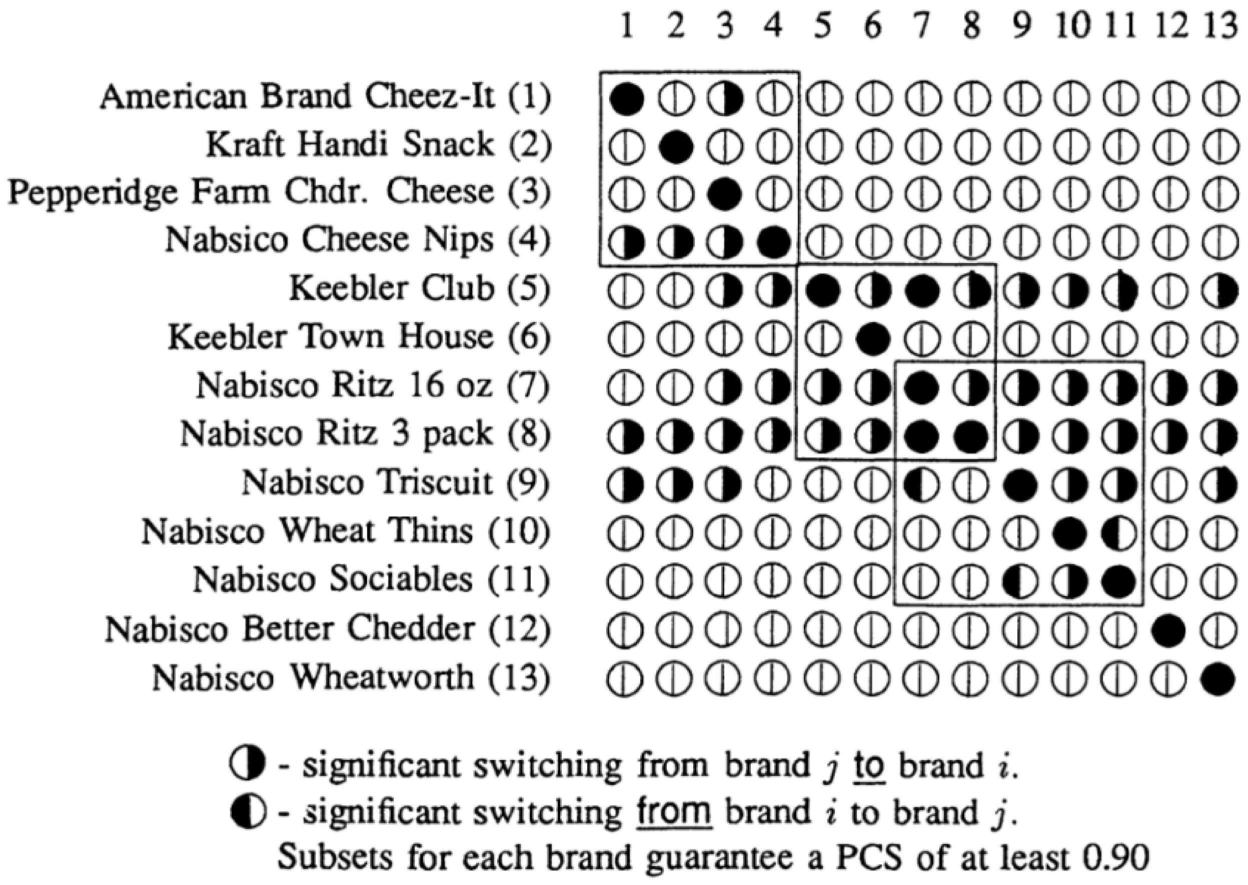
HIERARCHICAL STRUCTURE FOR SOFT DRINKS



[a]C = Cola, L = Lemon/Lime, and G = Ginger ale.

Focus on a focal brand
(Subset selection methodology, Kannan and Sanchez 1994)

(b) Subset Identification Graphs



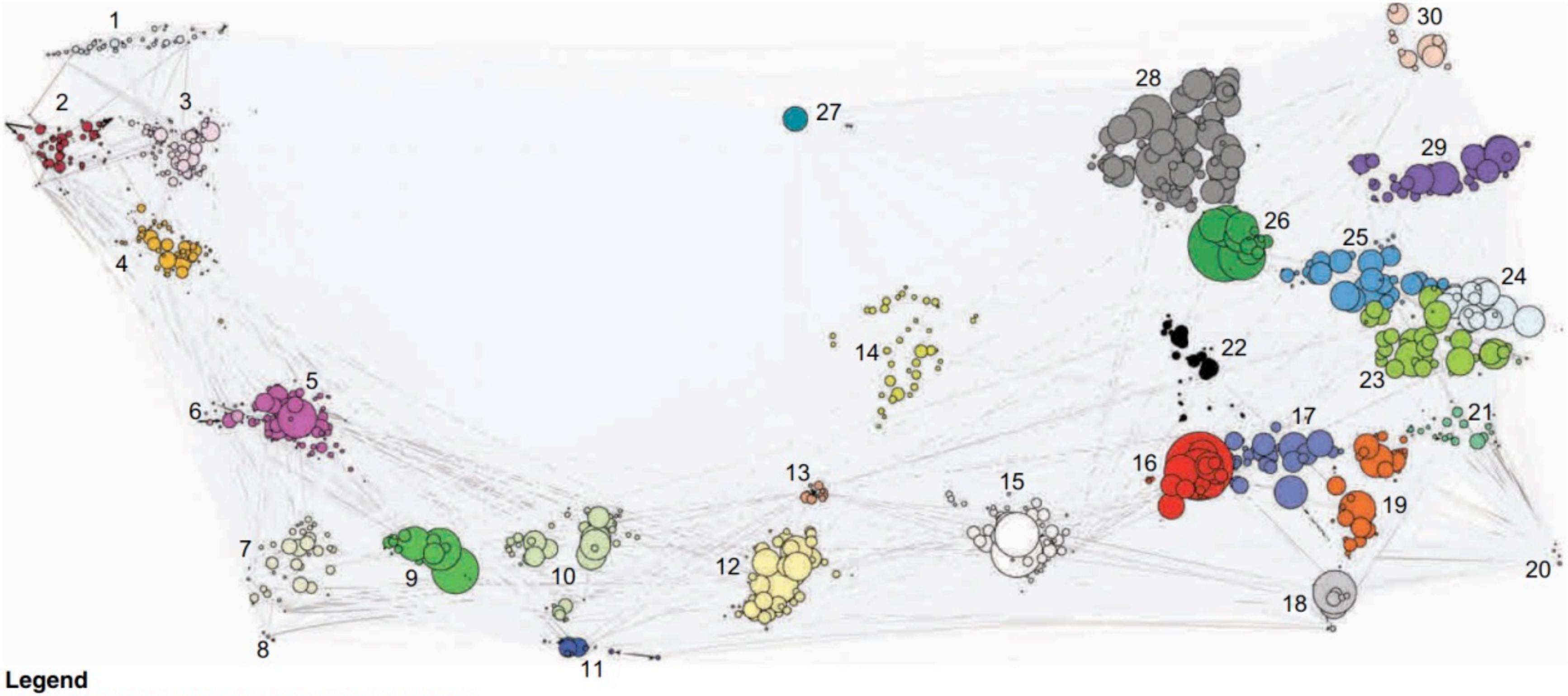
Evolution of literature

- Survey
  - Urban, Johnson and Hauser (1984)
  - Brand concept maps (BCM) (John et al. 2006)
  - ZMET (Zaltman and Coulter 1995)
- Scanner panel data
  - Grover and Srinivasan (1987)
  - Erdem (1996)
  - Lots of others…
- User click streams
  - e.g., Moe 2006
- Marketing mix
  - Carpenter and Lehmann (1985)
  - Kannan and Wright (1991)

Recent resurgence in big data context

(Search logs - Ringer and Skiera, MKS 2016), Online reviews - France and Ghose (MKS, 2016)

Figure 5 **Visualization of Asymmetric Competitive Market Structure Map of 1,124 LED-TVs**



Legend
Bubbles represent individual products (SKUs)
Bubble color indicates submarket membership
Bubble size indicates global competitive asymmetry (consideration frequency)
Arrows represent local competitive asymmetry and point at competitors of the product they originate in
Arrow weight indicates how intense a competitive relationship is: the darker and thicker the arrow, the more intense the relationship
Submarkets are numbered 1 through 30

Evolution of literature

- Online search logs
  - Kim, Albuquerque, and Bronnenberg 2011
  - Ringel and Skiera 2016
- User-generated content
  - Customer reviews (Lee and Bradlow 2011)
  - Forum discussions (Netzer et al. 2012)
  - Chatter (Tirunillai and Tellis 2014)
  - Hashtags (Nam, Joshi, and Kannan 2017)
- Store-level sales data
  - Gabel, Guhl, and Klapper 2019

	Primary/Survey Data	Text Mining (UGC)	Social Tag-based	Search Data	Social Engagement
Data Volume	Small	Large	Large	Large	Very large
Data Veracity	Authentic	Noisy	Moderate noisy	Moderate noisy	Moderate noisy
Privacy preserve	Yes	Yes	Yes	No (need to insert a tracking pixel)	Yes
Data availability	Low (need to collect data daily)	High (publicly available)	High (publicly available)	Low (need to insert a tracking pixel)	High (publicly available)
Data pre-processing cost	Low (use consideration set directly)	High (text mining is error-prone)	High (text mining is error-prone)	Low (use consideration set directly)	Low (use network raw data)

Comparison of different types of data

Differences among extant literature

	Kim et.al 2011	Lee and Bradlow 2011	Netzer et.al 2012	Ringel and Skiera 2016	Culotta and Cutler 2016	Nam, Joshi and Kannan 2017	Our study
Objective	To visualize user search behavior and understand market structure	To visualize competitive market structure using text mining on customer review	To visualize competitive market structure using text mining on forum discussion	To understand asymmetric competition in the product categories	To infer attribute-specific brand ratings	To analyze user generated tags for marketing research	To propose a novel deep network representation learning framework for marketing research
Brands/Products	62 products, 4 brands	9 brands	169 products, 30 brands	1,124 products	200 brands	7 brands	5,478 brands
Consumers/Users	N.A.	N.A.	76,587	100,000+	14.6 million	N.A.	25,992,832
Data sources	Amazon	Customer review at Epinions	Online discussion forum	Product comparison website	Twitter	Social tagging platform Delicious	Facebook public fan pages
Data type	Consumer search	Text	Text	Consumer search	Network	Social tags	Network
Brand association methodology	Consideration set	Text-mining	Text-mining	Consideration set	Network learning	Network learning	Network learning
Asymmetry	Yes	No	No	Yes	No	No	Yes
Dynamic	No	No	No	No	No	Yes	Yes
Dimension reduction	Yes	Yes	No	No	No	Yes	Yes
External validation	N.A.	N.A.	Purchase data, survey	Survey	Survey	Brand concept map (survey)	Event study, link prediction
Privacy preserve	Yes	Yes	Yes	No (need to insert a tracking pixel)	Yes	Yes	Yes
Data availability	Low (need to collect data daily)	High (publicly available)	High (publicly available)	Low (need to insert a tracking pixel)	High (publicly available)	High(publicly available)	High(publicly available)
Data preprocessing cost	Low (use consideration set directly)	High (text mining is error-prone)	High (text mining is error-prone)	Low (use consideration set directly)	Low (use network raw data)	Low (tags are well defined)	Low (use network raw data)

Proposed methodology

- We expect to:
  - Handle large-scale (easy-to-obtain) data
  - Learn complex and implicit patterns from data
  - Identify (sub)markets **without pre-specifying** boundaries
  - Capture dynamic changes of market structure

Data

- From social media platforms – Facebook
  - “Likes”
  - “Comments”
  - “Sharing”

- Nature of the data
  - higher-level brand metrics as compared to SKU-level

“Liking” brands on Facebook

Close to 90% of users on Facebook say that they “Like” at least one brand on Facebook (Lab42 survey)

50% say that they find the brand’s Facebook page more useful than the company’s website.

Of the Facebook users who “Like” brands:

- 82% said that Facebook is a good place to interact with brands
- 75% said that they felt more connected to the brand on Facebook
- 69% said that they Liked a brand because a friend in their network did

Why do they “like” the brands?

Reasons for Becoming a Brand Fan on Facebook

QUESTION: The following are the reasons of becoming a fan that were mentioned to us by others. Which, if any, of the following reasons led you to become a Fan or ‘Like’ the following brands on Facebook?

49%	To support the brand I like	27%	To share my interests / lifestyle with others
42%	To get a coupon or discount	21%	To research brands when I was looking for specific products / services
41%	To receive regular updates from brands I like	20%	Seeing my friends are already a fan or "liked"
35%	To participate in contests	18%	A brand advertisement (TV, online, magazines) led me to fan the brand
31%	To share my personal good experiences	15%	Someone recommended me to fan the brand

Syncapse/Hotspex U.S. Survey March 2013 (n=2,080). Primary brands under study included BMW, BlackBerry, Xbox , Disney,Zara, Levi's, H&M, Victoria's Secret, Adidas Originals, Nike, Monster Energy Drink, Coca-Cola, Dr Pepper , Oreo, Skittles, Starbucks, McDonald's, Subway, Walmart, Target.

Source: Syncapse.com

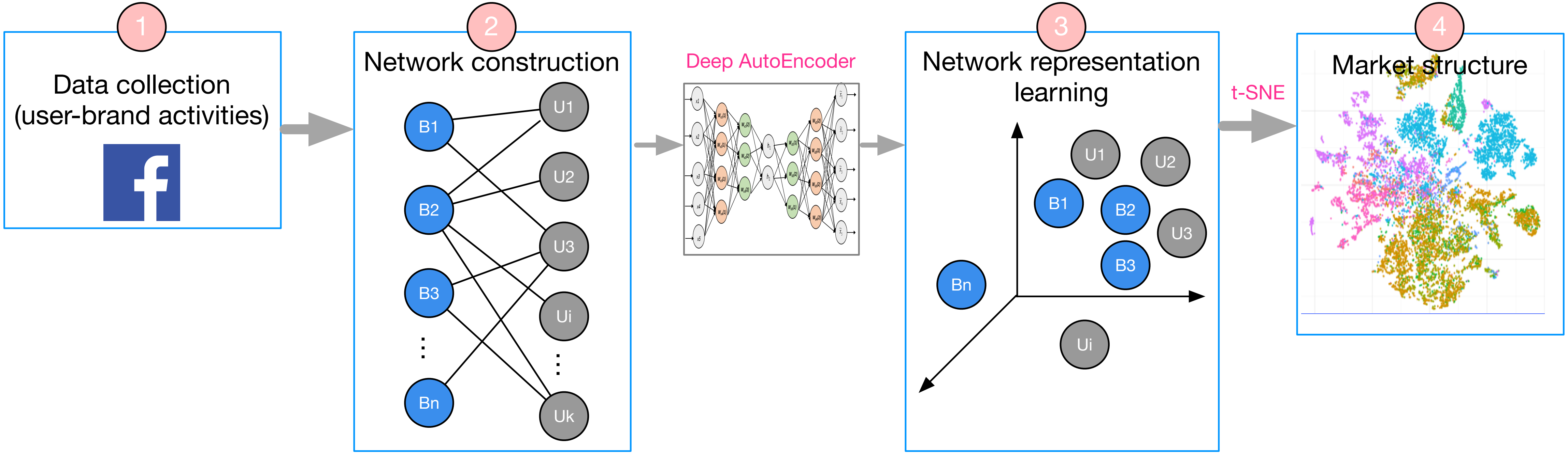
Syncapse

Does like translate to purchase? loyalty?

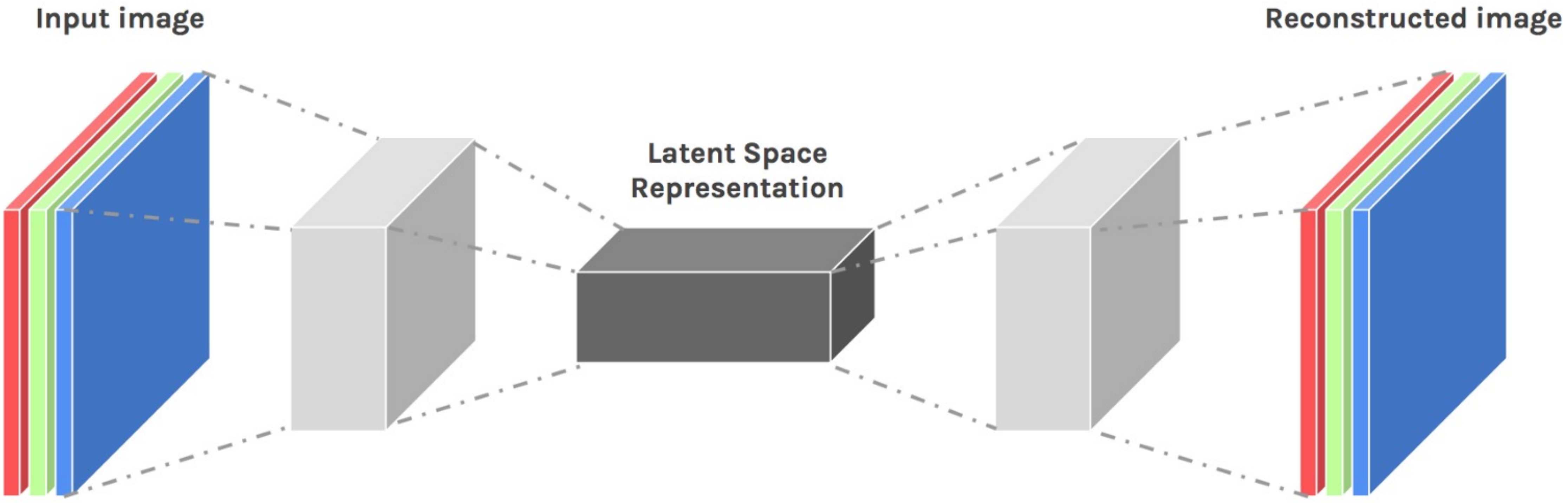
- **What Are Likes Worth? A Facebook Page Field Experiment (2017)**
  - Daniel Mochon, Karen Johnson, Janet Schwartz, Dan Ariely
- **Does "Liking" Lead to Loving? The Impact of Joining a Brand's Social Network on Marketing Outcomes (2017)**
  - Leslie K. John, Oliver Emrich, Sunil Gupta, Michael I. Norton

- We are more interested in the information on content, user engagement with brands

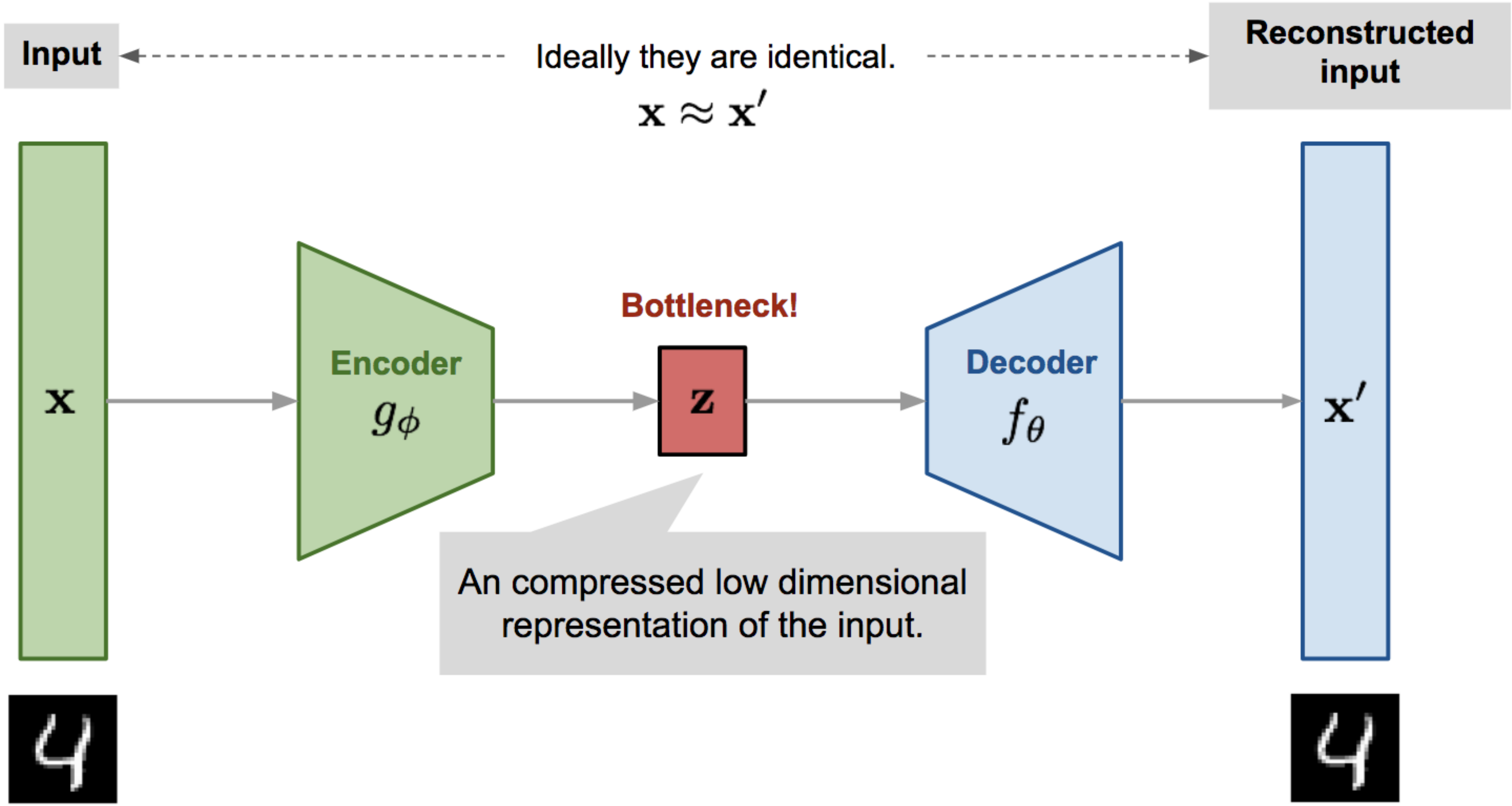
Our proposed approach – overall framework



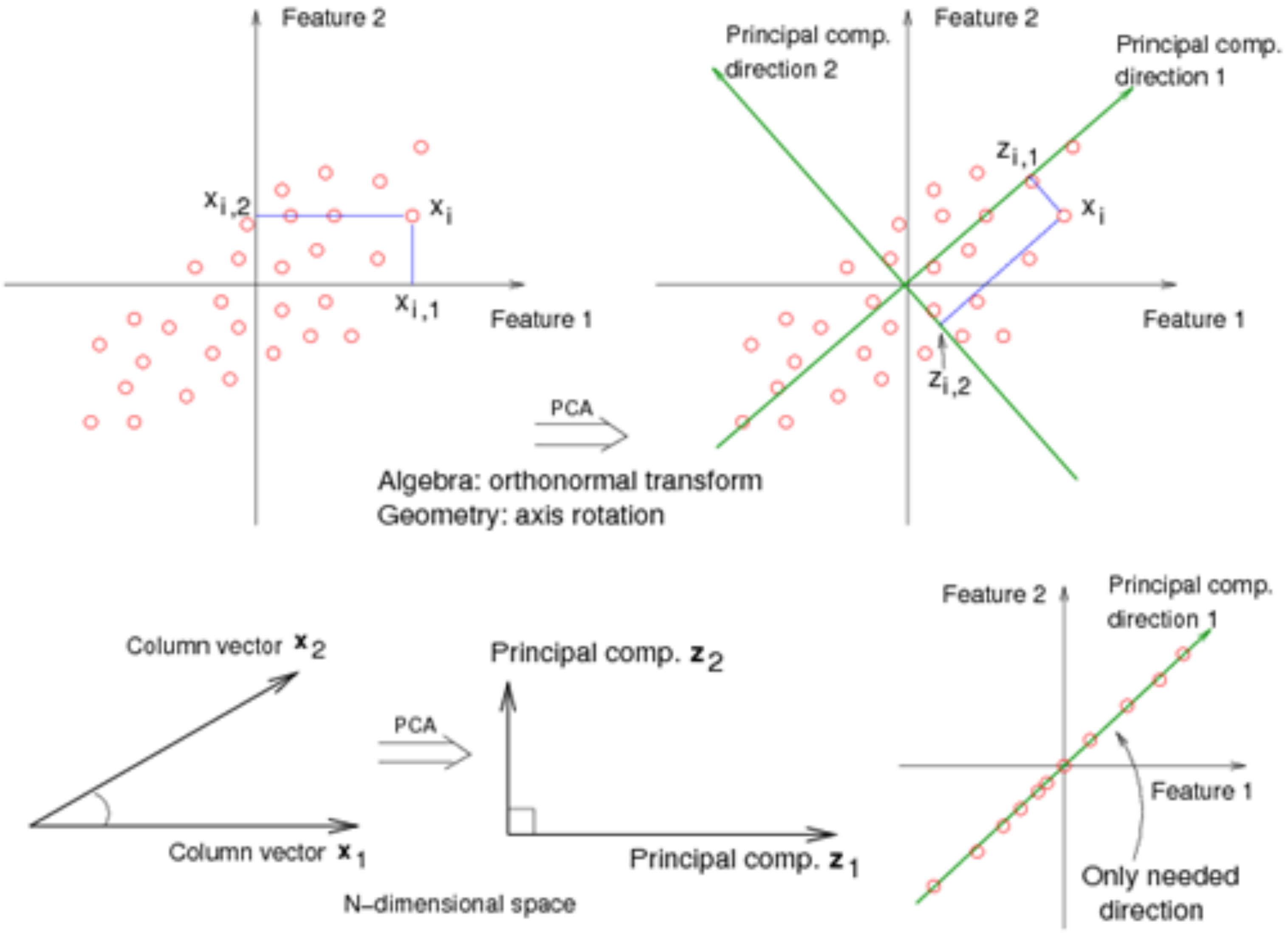
Deep autoencoders



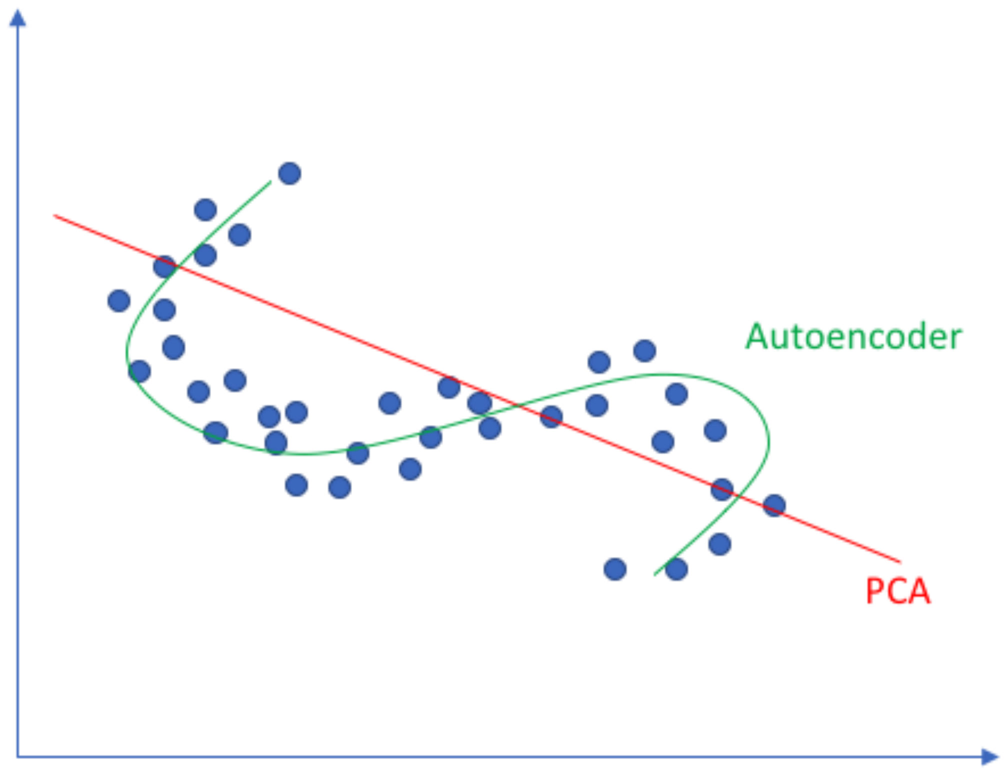
Deep autoencoders



PCA


Feature 2
Feature 1
$x_{i,2}$
x_i
$x_{i,1}$
PCA
Algebra: orthonormal transform
Geometry: axis rotation
Principal comp. direction 2
Principal comp. direction 1
$z_{i,1}$
$z_{i,2}$
Column vector $\mathbf{x}_2$
Column vector $\mathbf{x}_1$
Principal comp. $\mathbf{z}_2$
Principal comp. $\mathbf{z}_1$
N-dimensional space
Only needed direction

Linear vs nonlinear dimensionality reduction



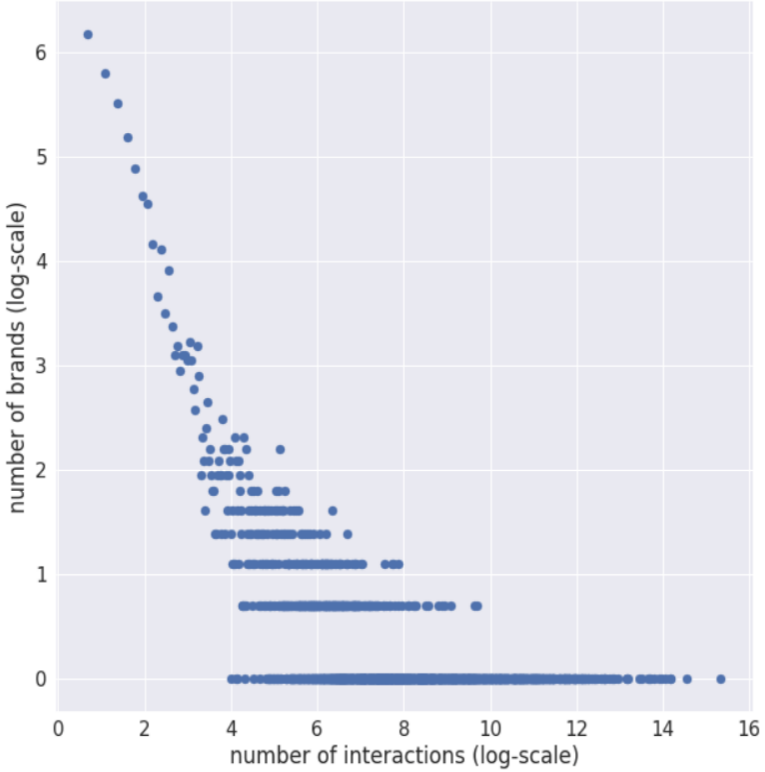
Data collection

- Facebook public pages
  - Top list of US brands based on #followers from Socialbakers.com
  - 25 different categories: **brands (our focus)**, celebrities, community, entertainment, media, places, society and sport, etc.
  - Graph API to collect all user-brand interactions: posts, comments, likes, and shares.
  - Jan. 1, 2017 – Jan. 1, 2018 for analysis

Number of brands	5,478
Number of users	25,992,832
Number of unique user-brand interactions	36,927,613
Number of like interactions	87,876,623
Number of unique user-brand like interactions	29,611,805
Number of comment interactions	18,703,549
Number of unique user-brand comment interactions	7,612,358
Total number of user-brand interactions	106,580,172

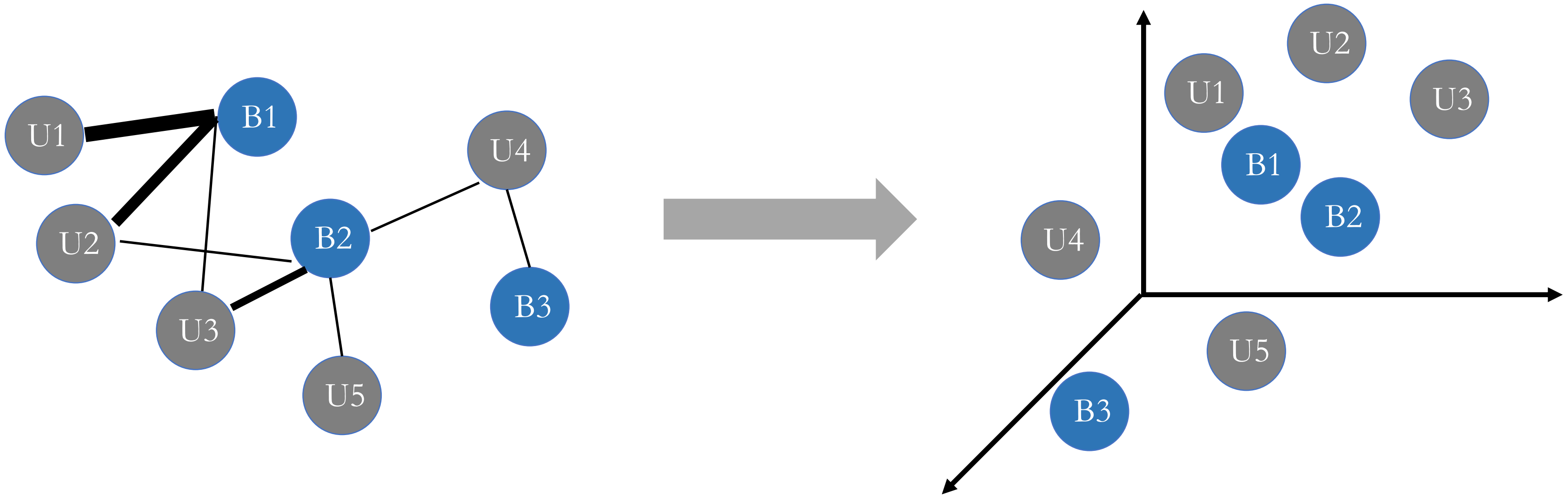
Data collection

- Data cleansing
  - Fake user removal (simple but effective rules following previous works) (Zhang et al. 2016)



3. Deep network representation learning

- Mathematically, given a large information network, our method aims to learn node representations in a low dimensional space



- Learning objective: preserve local/global network structures and semantics as much as possible
  - Minimize the total loss: $L_{1st} + L_{2nd}$ and the reconstruction error: e_{rec}

3.1 First order similarity

- Similarity to neighbors
  - The local pairwise similarity between user node and brand node
  - The edge weight indicates the similarity strength between two nodes.
    - If there is no edge between two nodes, their first-order similarity is almost 0

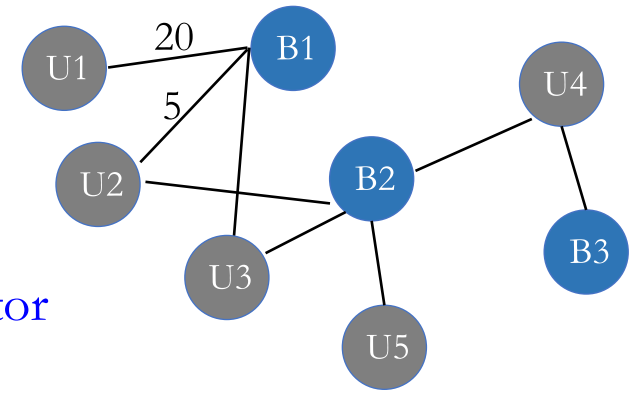
edge weight

$$L_{1st} = \sum_{i=1}^{m}\sum_{j=1}^{n} e_{ij}\left(w_i^b - w_j^u\right)^2$$

Brand representation vector in the learned space

User representation vector in the learned space

i.e., output of encoder



3.2 Second order similarity

- Similarity to neighbors of neighbors
  - The similarity of a node with its neighbor’s neighbor, such as brand node and another brand node; user node and another user node
    - If two nodes do not have any intermediate nodes in between, their second-order similarity is close to 0

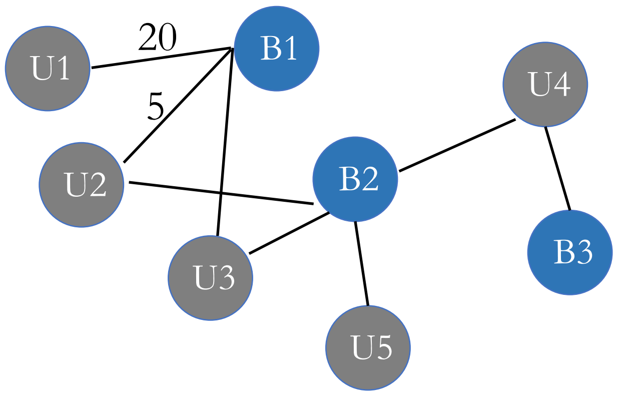
Number of shared users

Number of shared brands

$$L_{2nd} = \sum_{i=1}^{m}\sum_{j=1}^{m} c_{ij}^{u}\left(w_i^b - w_j^b\right)^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}^{b}\left(w_i^u - w_j^u\right)^2$$

Brand representation vector in the learned space

User representation vector in the learned space



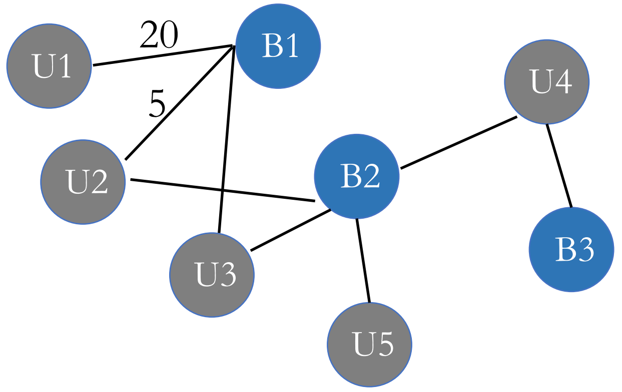
3.3 Reconstruction error

- Minimize the reconstruction error between the learned representation and the original representation

Autoencoder input: user and brand vector representation using one-hot encoding

$$e_{rec} = \sum_{i=1}^{m} \left(w_i^{b'} - w_i^{b}\right)^2 + \sum_{j=1}^{n} \left(w_j^{u'} - w_j^{u}\right)^2$$

Autoencoder output: reconstructed user and brand vector representation



4. Market structure discovery

- The output of the *K*-th layer (last layer of encoder) is the learned representation (e.g., 300 dimensional vector) for market structure discovery

- Further dimension reduction for visualization
    - t-Distributed Stochastic Neighbor Embedding (t-SNE) (L.J.P. van der Maaten, 2014)

Evaluation

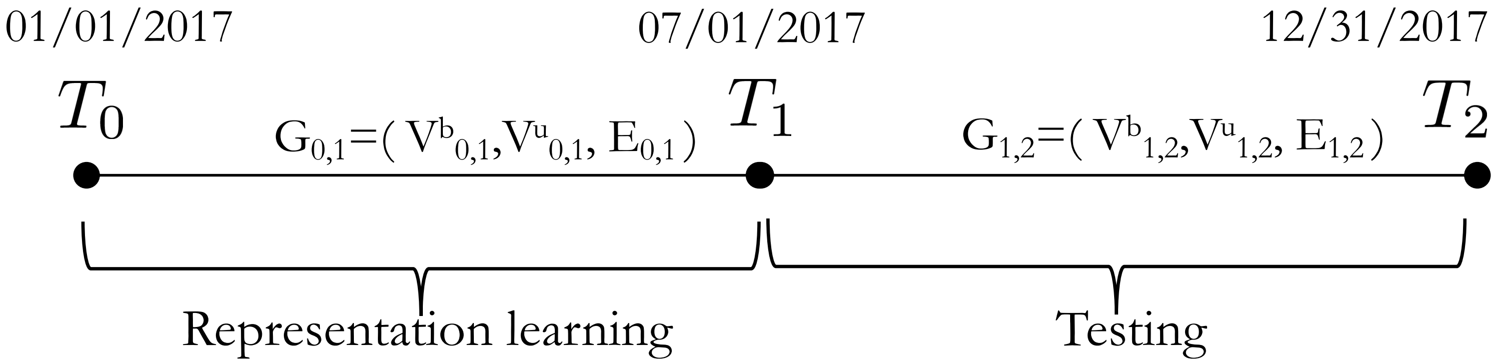
- Challenges
  - Lack of ground truth for market structure discovery
  - Industry classification (e.g., SIC or NAICS)
    - Static - do not re-classify firms over time

- Key: brand representation

- Alternative evaluation: link prediction
  - Good representation: should well capture latent, complicated semantic, and structural information among brands.

    Naylor, Lamberton, and West 2012; Kuksov, Shachar, and Wang 2013; Culotta and Cutler 2016

Link prediction



- Algorithm (input: $G_{0,1}$ and $G_{1,2}$)
  1. Learn low-dimensional representation for each user and brand in the training period;
  2. Randomly select N users (e.g., N=100, N=1000);
  3. Initialize an empty set $S = \Phi$;
  4. Foreach user u_i in N users:

     Foreach brand b_j in all existing brands, do:

     Calculate the proximity score between u_i and b_j: s_{ij};

     $S \leftarrow (u_i, b_j, s_{ij})$;
  5. End For
  6. Sort S w.r.t. s_{ij} to get top n user-brand pairs (denoted as P);
  7. Calculate precision@n and recall@n: $precision@n = \frac{|P \cap E_{1,2}|}{n}$, $recall@n = \frac{|P \cap E_{1,2}|}{|E^T_{1,2}|}$

The set of all newly formed links in $G_{1,2}$ for brands and users appeared in the training period

Link prediction

- Baselines and variants
  - 2 X 2 design

Network	Homogeneous	Brand-brand network derived from the original user-brand network (Zhang et al. 2016; Culotta and Cutler 2016; etc.)
	Heterogeneous	The original user-brand network (preserve semantics)
Model	Shallow	Matrix factorization (user-brand matrix) (latent representation – not deep, ignore structural information)
	Deep	Our deep Autoencoder representation learning (capture deep structures and semantics encoded in the network)

Confusion Matrix	Positive (Predicted)	Negative (Predicted)
Positive (Actual)	True Positive (TP)	False Negative (FN)
Negative (Actual)	False Positive (FP)	True Negative (TN)

Recall = TP/(TP +FN)

Precision = TP/(TP + FP)

F1 = 2*Precision*Recall/(Precision + Recall)

Link prediction results

precision@n		n=10	n=100	n=500	n=1000	n=5,000	n=10,000	n=100,000
Homogeneous brand-brand network	Shallow model	0.400	0.262	0.132	0.078	0.022	0.012	0.001
		(0.109)	(0.023)	(0.018)	(0.008)	(0.002)	(0.000)	(0.000)
	Deep model	0.410	0.271	0.139	0.082	0.023	0.014	0.001
		(0.092)	(0.027)	(0.020)	(0.009)	(0.003)	(0.001)	(0.000)
Heterogenous brand-user network	Shallow model	0.430	0.291	0.157	0.095	0.028	0.018	0.001
		(0.102)	(0.030)	(0.024)	(0.008)	(0.005)	(0.002)	(0.000)
	Deep model	**0.52*****	**0.322****	**0.173****	**0.124*****	**0.034*****	**0.028*****	**0.001*****
		(0.092)	(0.022)	(0.051)	(0.011)	(0.008)	(0.001)	(0.000)

- The number of randomly selected users: 100

Link prediction results

recall@n		n=10	n=100	n=500	n=1000	n=5,000	n=10,000	n=100,000
Homogeneous brand-brand network	Shallow model	0.031	0.260	0.488	0.602	0.828	0.918	0.996
		(0.008)	(0.002)	(0.060)	(0.050)	(0.036)	(0.016)	(0.005)
	Deep model	0.032	0.275	0.505	0.621	0.832	0.912	0.997
		(0.013)	(0.032)	(0.054)	(0.047)	(0.049)	(0.032)	(0.003)
Heterogenous brand-user network	Shallow model	0.037	0.287	0.521	0.637	0.870	0.935	0.998
		(0.015)	(0.065)	(0.074)	(0.045)	(0.023)	(0.047)	(0.000)
	Deep model	**0.056****	**0.311****	**0.582****	**0.686****	**0.897****	**0.967****	**0.999****
		(0.013)	(0.035)	(0.077)	(0.054)	(0.078)	(0.024)	(0.002)

- The number of randomly selected users: 100

Link prediction results

precision@n		n=10	n=100	n=500	n=1000	n=5,000	n=10,000	n=100,000
Homogeneous brand-brand network	Shallow model	0.460	0.387	0.331	0.291	0.130	0.078	0.012
		(0.132)	(0.112)	(0.021)	(0.012)	(0.004)	(0.003)	(0.000)
	Deep model	0.490	0.393	0.332	0.295	0.131	0.078	0.012
		(0.020)	(0.003)	(0.018)	(0.017)	(0.003)	(0.003)	(0.000)
Heterogenous brand-user network	Shallow model	0.500	0.422	0.344	0.320	0.162	0.087	0.012
		(0.102)	(0.060)	(0.022)	(0.072)	(0.010)	(0.017)	(0.000)
	Deep model	**0.522*** **	**0.436*** **	**0.365*** **	**0.355*** **	**0.187*** **	**0.091*** **	**0.013*** **
		(0.092)	(0.040)	(0.012)	(0.035)	(0.014)	(0.047)	(0.000)

- The number of randomly selected users: 1,000

Link prediction results

recall@n		n=10	n=100	n=500	n=1000	n=5,000	n=10,000	n=100,000
Homogeneous brand-brand network	Shallow model	0.031	0.033	0.128	0.223	0.509	0.607	0.915
		(0.008)	(0.021)	(0.008)	(0.008)	(0.013)	(0.013)	(0.008)
	Deep model	0.032	0.035	0.131	0.226	0.510	0.605	0.921
		(0.005)	(0.047)	(0.018)	(0.011)	(0.010)	(0.015)	(0.007)
Heterogenous brand-user network	Shallow model	0.049	0.056	0.365	0.241	0.549	0.658	0.981
		(0.022)	(0.009)	(0.012)	(0.010)	(0.012)	(0.024)	(0.015)
	Deep model	**0.049***************	**0.076***************	**0.412***************	**0.352***************	**0.584***************	**0.743*****	**0.990***************
		(0.009)	(0.003)	(0.010)	(0.007)	(0.009)	(0.008)	(0.002)

- The number of randomly selected users: 1,000

Link prediction results

precision@1000		10%	30%	50%	70%	90%	100%
Homogeneous brand-brand network	Shallow model	0.103	0.195	0.248	0.263	0.282	0.291
		(0.012)	(0.008)	(0.008)	(0.012)	(0.015)	(0.012)
	Deep model	0.097	0.190	0.248	0.267	0.284	0.295
		(0.042)	(0.010)	(0.021)	(0.031)	(0.023)	(0.017)
Heterogenous brand-user network	Shallow model	0.143	0.225	0.256	0.283	0.312	0.320
		(0.015)	(0.031)	(0.042)	(0.008)	(0.052)	(0.072)
	Deep model	**0.183*** **	**0.242*** **	**0.273*** **	**0.301*** **	**0.337*** **	**0.355*** **
		(0.024)	(0.032)	(0.037)	(0.012)	(0.032)	(0.035)

- The number of randomly selected users: 1,000

Impact of training size

precision@1000		10%	30%	50%	70%	90%	100%
Homogeneous brand-brand network	Shallow model	0.103	0.195	0.248	0.263	0.282	0.291
		(0.012)	(0.008)	(0.008)	(0.012)	(0.015)	(0.012)
	Deep model	0.097	0.190	0.248	0.267	0.284	0.295
		(0.042)	(0.010)	(0.021)	(0.031)	(0.023)	(0.017)
Heterogenous brand-user network	Shallow model	0.143	0.225	0.256	0.283	0.312	0.320
		(0.015)	(0.031)	(0.042)	(0.008)	(0.052)	(0.072)
	Deep model	**0.183*** **	**0.242*** **	**0.273*** **	**0.301*** **	**0.337*** **	**0.355*** **
		(0.024)	(0.032)	(0.037)	(0.012)	(0.032)	(0.035)

- The number of randomly selected users: 1,000

Impact of training size

recall@1000		10%	30%	50%	70%	90%	100%
Homogeneous brand-brand network	Shallow model	0.080	0.153	0.193	0.203	0.219	0.223
		(0.009)	(0.006)	(0.006)	(0.007)	(0.011)	(0.008)
	Deep model	0.075	0.150	0.194	0.204	0.220	0.226
		(0.005)	(0.010)	(0.007)	(0.003)	(0.005)	(0.011)
Heterogenous brand-user network	Shallow model	0.108	0.179	0.223	0.257	0.271	0.241
		(0.031)	(0.018)	(0.013)	(0.026)	(0.017)	(0.010)
	Deep model	**0.124*****	**0.198*****	**0.24*****	**0.289*****	**0.314*****	**0.352*****
		(0.009)	(0.008)	(0.019)	(0.029)	(0.008)	(0.007)

- The number of randomly selected users: 1000

Like network only

- The number of randomly selected users: 1,000

precision@n		n=10	n=100	n=500	n=1,000	n=5,000	n=10,000	n=100,000
Homogeneous brand-brand network	Linear model	0.320	0.279	0.258	0.233	0.127	0.067	0.011
		(0.094)	(0.056)	(0.008)	(0.008)	(0.004)	(0.001)	(0.001)
	Deep model	0.323	0.284	0.258	0.235	0.135	0.069	0.011
		(0.147)	(0.082)	(0.017)	(0.009)	(0.014)	(0.034)	(0.002)
Heterogenous brand-user network	Linear model	0.424	0.365	0.312	0.287	0.152	0.087	0.011
		(0.035)	(0.042)	(0.039)	(0.008)	(0.032)	(0.003)	(0.000)
	Deep model	**0.486******	**0.398******	**0.354******	**0.314******	**0.178******	**0.091******	**0.011**
		(0.026)	(0.032)	(0.023)	(0.009)	(0.037)	(0.004)	(0.001)

recall@n		n=10	n=100	n=500	n=1,000	n=5,000	n=10,000	n=100,000
Homogeneous brand-brand network	Linear model	0.002	0.024	0.111	0.201	0.458	0.563	0.896
		(0.001)	(0.005)	(0.003)	(0.006)	(0.015)	(0.010)	(0.006)
	Deep model	0.002	0.025	0.124	0.204	0.476	0.560	0.882
		(0.002)	(0.002)	(0.011)	(0.018)	(0.052)	(0.023)	(0.034)
Heterogenous brand-user network	Linear model	0.041	0.056	0.332	0.350	0.521	0.635	0.911
		(0.003)	(0.004)	(0.029)	(0.029)	(0.075)	(0.079)	(0.009)
	Deep model	**0.049*****	**0.068*****	**0.350*****	**0.404*****	**0.562*****	**0.663*****	**0.929*****
		(0.005)	(0.006)	(0.021)	(0.043)	(0.037)	(0.063)	(0.028)

Comment network only

- The number of randomly selected users: 1,000

precision@n		n=10	n=100	n=500	n=1,000	n=5,000	n=10,000	n=100,000
Homogeneous brand-brand network	Linear model	0.189	0.179	0.156	0.134	0.067	0.045	0.010
		(0.169)	(0.041)	(0.014)	(0.008)	(0.005)	(0.003)	(0.000)
	Deep model	0.189	0.168	0.162	0.137	0.062	0.044	0.010
		(0.097)	(0.019)	(0.052)	(0.010)	(0.032)	(0.002)	(0.001)
Heterogenous brand-user network	Linear model	0.213	0.192	0.167	0.154	0.122	0.080	0.010
		(0.025)	(0.087)	(0.029)	(0.024)	(0.052)	(0.020)	(0.001)
	Deep model	**0.234*** **	**0.210*** **	**0.173*** **	**0.168*** **	**0.126*** **	**0.088*** **	**0.011* **
		(0.045)	(0.023)	(0.067)	(0.019)	(0.033)	(0.002)	(0.002)

recall@n		n=10	n=100	n=500	n=1,000	n=5,000	n=10,000	n=100,000
Homogeneous brand-brand network	Linear model	0.002	0.017	0.068	0.117	0.291	0.393	0.834
		(0.002)	(0.003)	(0.006)	(0.008)	(0.017)	(0.018)	(0.008)
	Deep model	0.002	0.019	0.068	0.114	0.295	0.393	0.842
		(0.001)	(0.012)	(0.022)	(0.032)	(0.042)	(0.053)	(0.012)
Heterogenous brand-user network	Linear model	0.019	0.042	0.077	0.162	0.333	0.442	0.885
		(0.003)	(0.019)	(0.045)	(0.029)	(0.029)	(0.056)	(0.034)
	Deep model	**0.018**	**0.044** **	**0.082*** **	**0.182*** **	**0.352*** **	**0.453*** **	**0.894*** **
		(0.004)	(0.012)	(0.051)	(0.037)	(0.026)	(0.033)	(0.046)

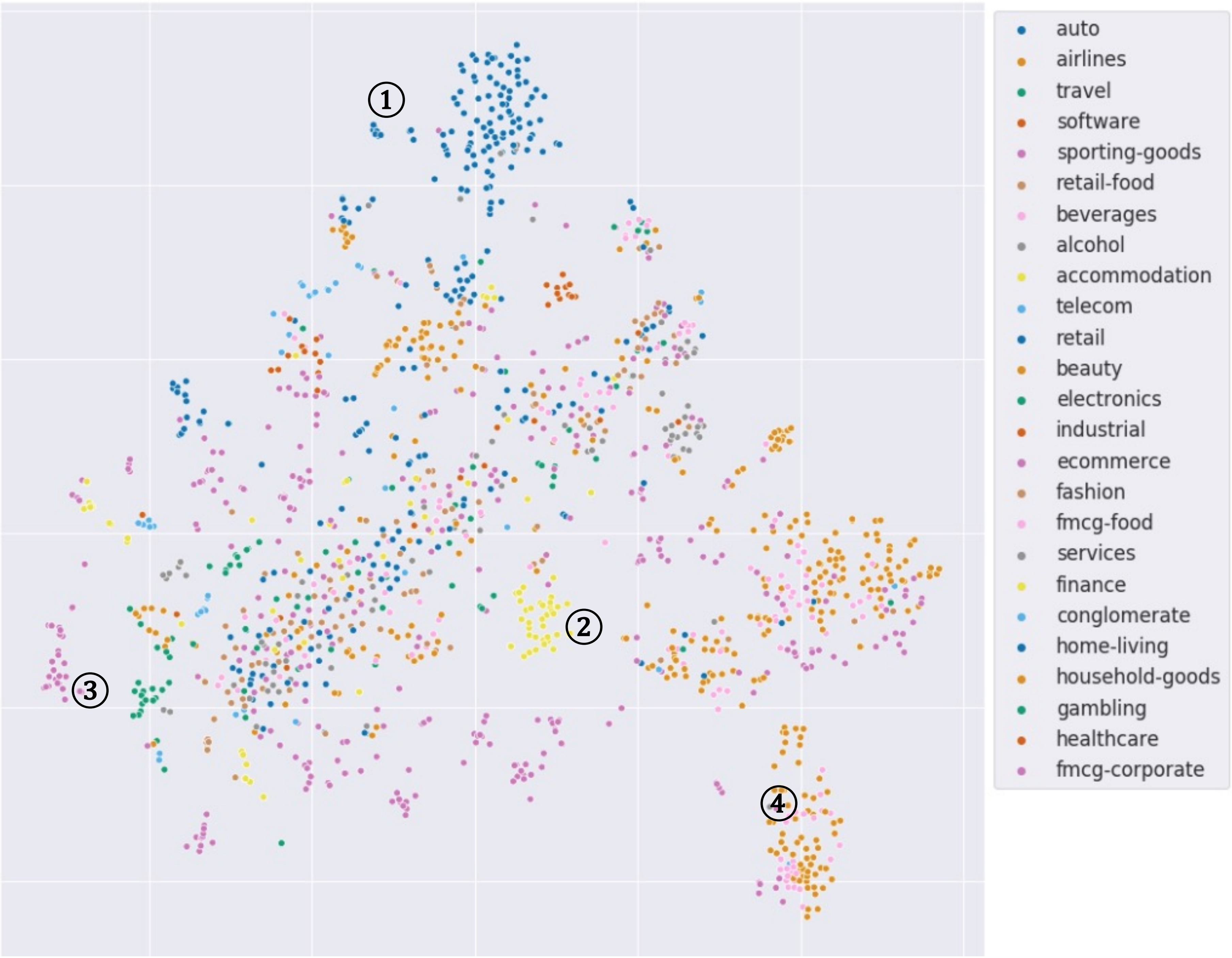
Extra validation

- Amazon Mechanical Turk (AMT)
  - 28 auto brands
  - 28x28 survey matrix – brand-brand similarity
  - 28x28 deep-learning matrix
  - Correlation is significantly positive ($r = 0.385$, *p-value* = 0.000)

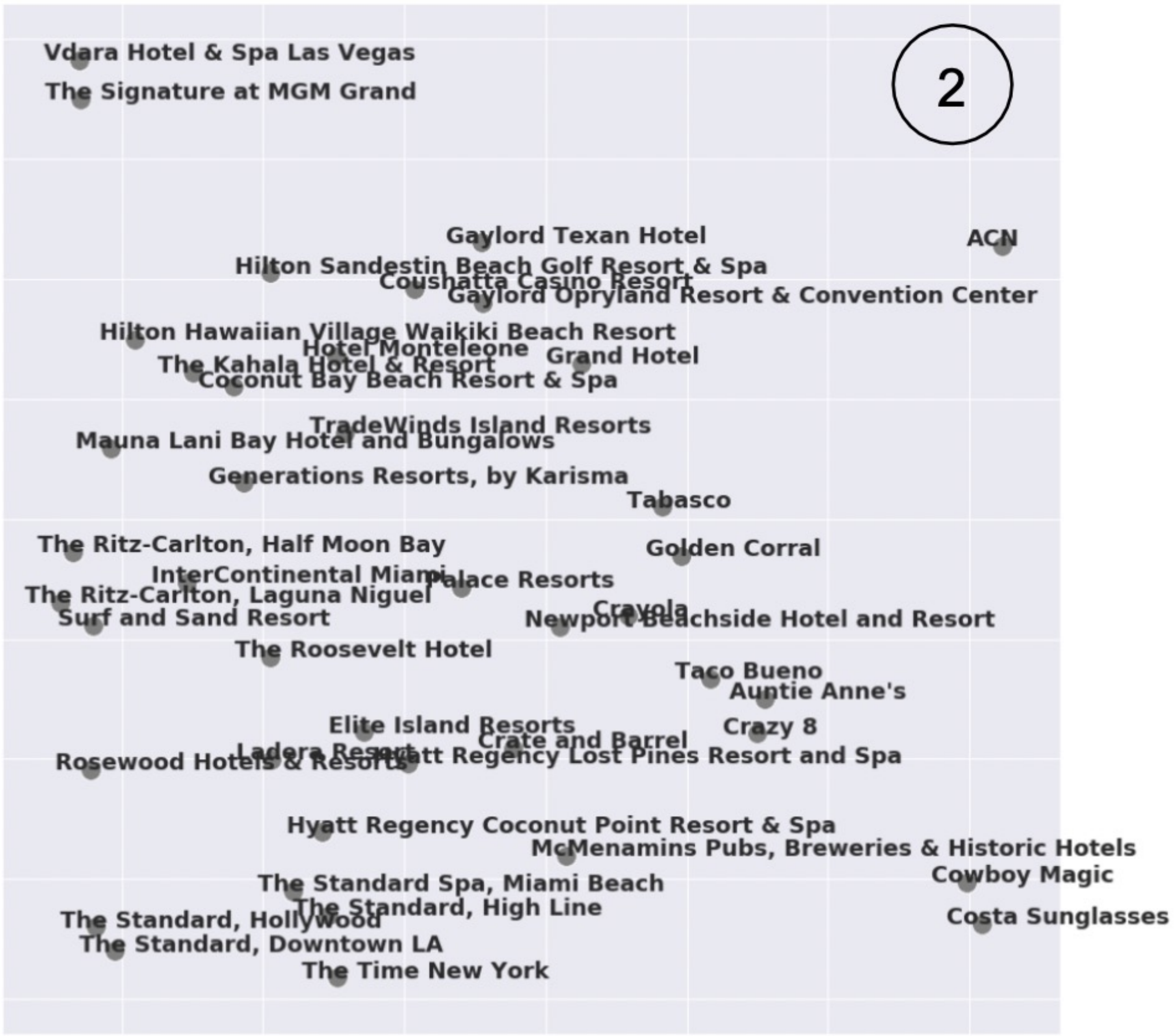
- Google search interest score
  - 19 airlines
  - Pearson's two-tailed correlation between two sets of 361 (=19*19) similarity scores
  - significantly and highly correlated ($r = 0.630$, *p-value* = 0.0000)

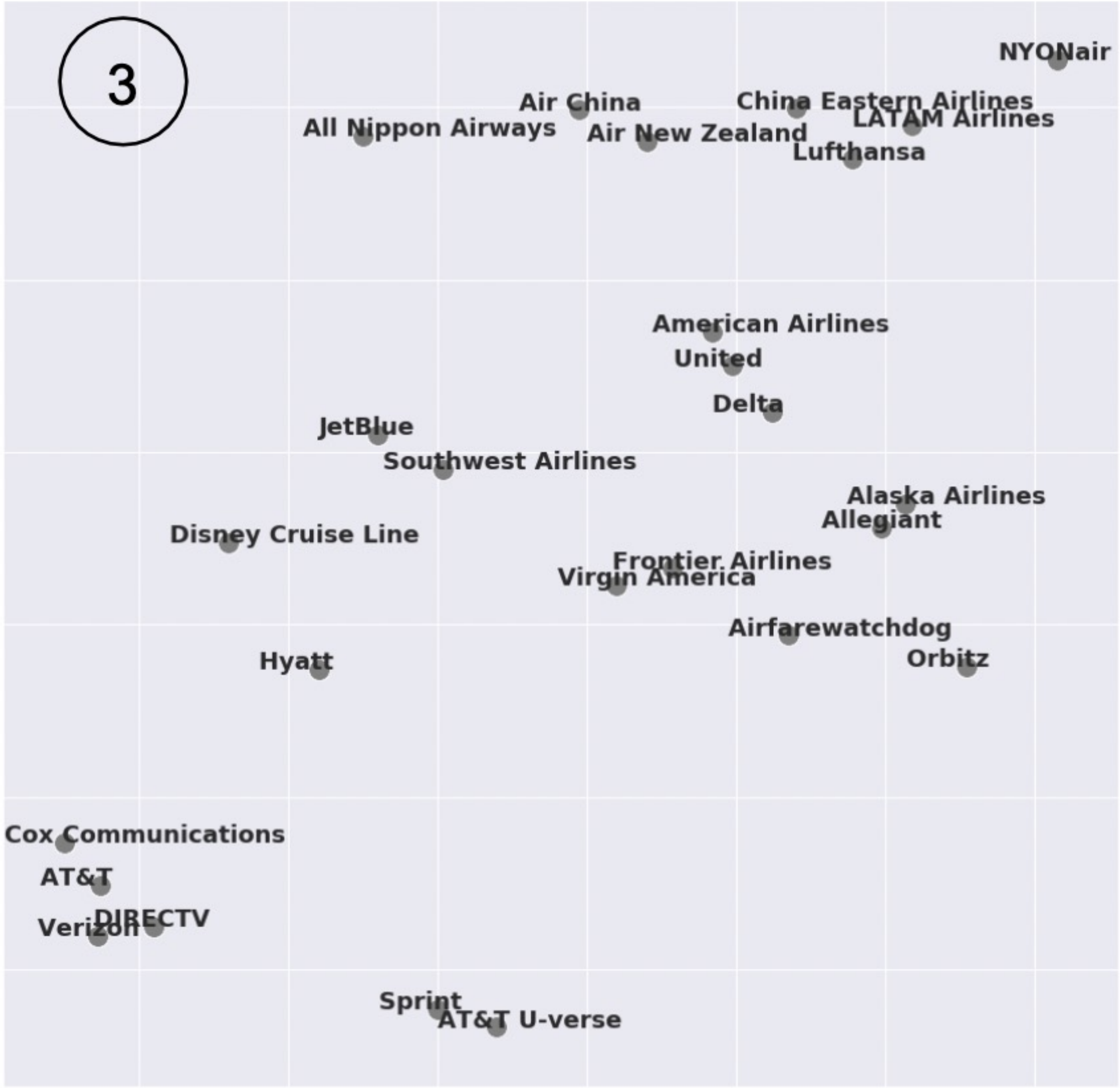
Global market structure visualization

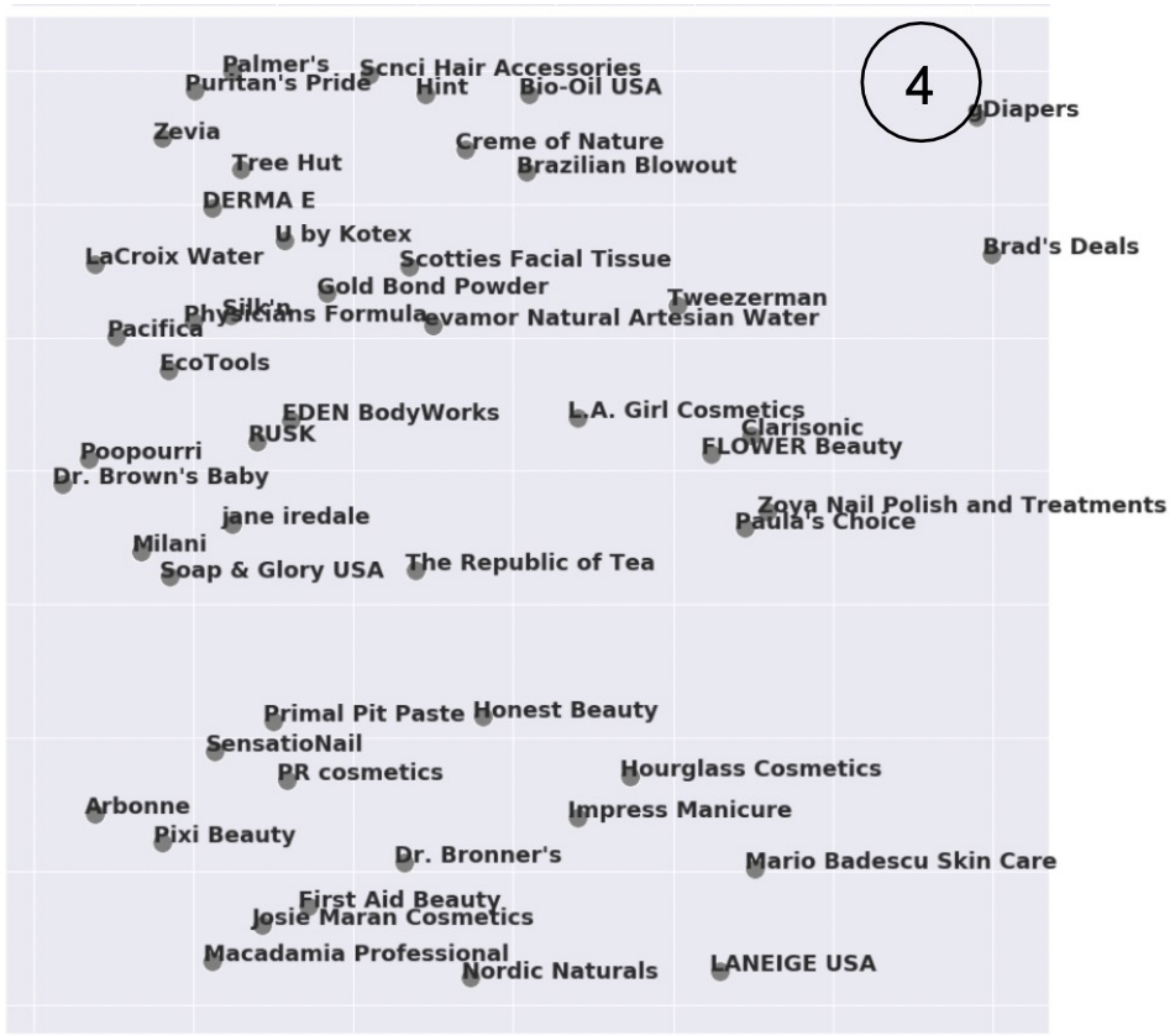
https://market-structure.github.io



Zoom-in on each cluster


2
Vdara Hotel & Spa Las Vegas
The Signature at MGM Grand
Gaylord Texan Hotel
ACN
Hilton Sandestin Beach Golf Resort & Spa
Coushatta Casino Resort
Gaylord Opryland Resort & Convention Center
Hilton Hawaiian Village Waikiki Beach Resort
Hotel Monteleone
Grand Hotel
The Kahala Hotel & Resort
Coconut Bay Beach Resort & Spa
TradeWinds Island Resorts
Mauna Lani Bay Hotel and Bungalows
Generations Resorts, by Karisma
Tabasco
The Ritz-Carlton, Half Moon Bay
Golden Corral
InterContinental Miami
Palace Resorts
The Ritz-Carlton, Laguna Niguel
Crayola
Surf and Sand Resort
Newport Beachside Hotel and Resort
The Roosevelt Hotel
Taco Bueno
Auntie Anne's
Elite Island Resorts
Crazy 8
Crate and Barrel
Ladera Resort
Hyatt Regency Lost Pines Resort and Spa
Rosewood Hotels & Resorts
Hyatt Regency Coconut Point Resort & Spa
McMenamins Pubs, Breweries & Historic Hotels
Cowboy Magic
The Standard Spa, Miami Beach
The Standard, High Line
The Standard, Hollywood
Costa Sunglasses
The Standard, Downtown LA
The Time New York

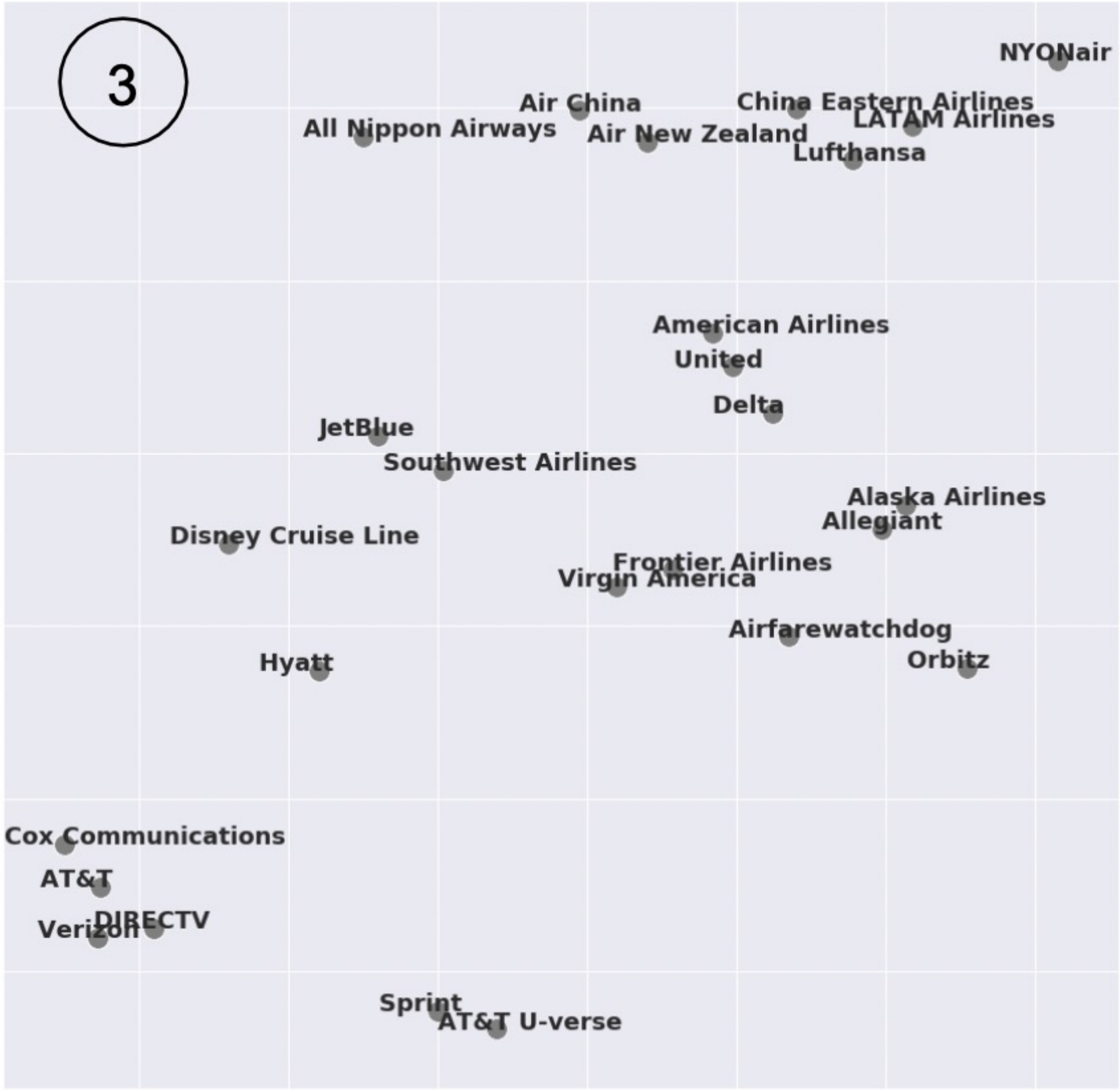

3
NYONair
Air China
China Eastern Airlines
LATAM Airlines
All Nippon Airways
Air New Zealand
Lufthansa
American Airlines
United
Delta
JetBlue
Southwest Airlines
Alaska Airlines
Allegiant
Disney Cruise Line
Frontier Airlines
Virgin America
Airfarewatchdog
Orbitz
Hyatt
Cox Communications
AT&T
DIRECTV
Verizon
Sprint
AT&T U-verse

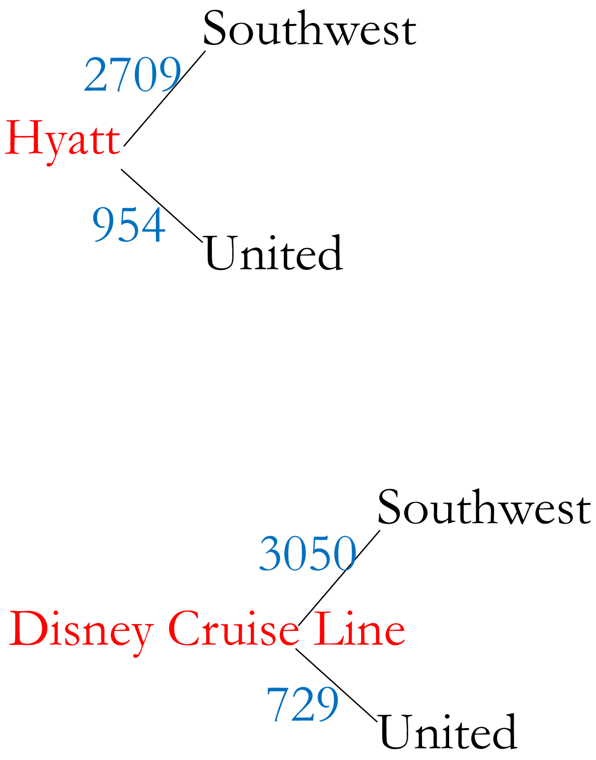

4
Palmer's
Scnci Hair Accessories
Puritan's Pride
Hint
Bio-Oil USA
gDiapers
Zevia
Creme of Nature
Tree Hut
Brazilian Blowout
DERMA E
U by Kotex
Brad's Deals
LaCroix Water
Scotties Facial Tissue
Gold Bond Powder
Tweezerman
Silk'n
Physicians Formula
evamor Natural Artesian Water
Pacifica
EcoTools
EDEN BodyWorks
L.A. Girl Cosmetics
Clarisonic
RUSK
FLOWER Beauty
Poopourri
Dr. Brown's Baby
Zoya Nail Polish and Treatments
jane iredale
Paula's Choice
Milani
The Republic of Tea
Soap & Glory USA
Primal Pit Paste
Honest Beauty
SensatioNail
Hourglass Cosmetics
PR cosmetics
Arbonne
Impress Manicure
Pixi Beauty
Dr. Bronner's
Mario Badescu Skin Care
First Aid Beauty
Josie Maran Cosmetics
Macadamia Professional
LANEIGE USA
Nordic Naturals

Identify similar brands

Focal brand		United	Southwest Airlines	Audi USA	Nissan
Rank	1	American	JetBlue	Mercedes-Benz USA	Mazda
	2	Delta	Frontier	BMW USA	Toyota
	3	Lufthansa	Allegiant	Land Rover	Volkswagen
	4	Southwest	Delta	Lexus	Kia Motors America
	5	Alaska	Alaska	Chevrolet Camaro	Subaru of America
	6	All Nippon	United	Maserati USA	Chrysler
	7	Air China	Airfarewatchdog	Kawasaki USA	FIAT
	8	LATAM	American	Firestone Tires	Jaguar
	9	Air New Zealand	Virgin America	Tesla	Alfa Romeo
	10	Airfarewatchdog	Hyatt	Ram Trucks	KLIM

Identify opportunities/threats





Small brands

Predominantly located in 2 areas

“The Luxury Travel Expert” - an information portal for luxury travel and premium tours, 11,000 followers as of data collection

Most similar brands: expert-led, small-group, luxury, and premium tours

“Smithsonian Journeys”
“The Peninsula Beverly Hills”
“Peter Sommer Travels”
“Quasar Expeditions”
“DuVine Cycling”

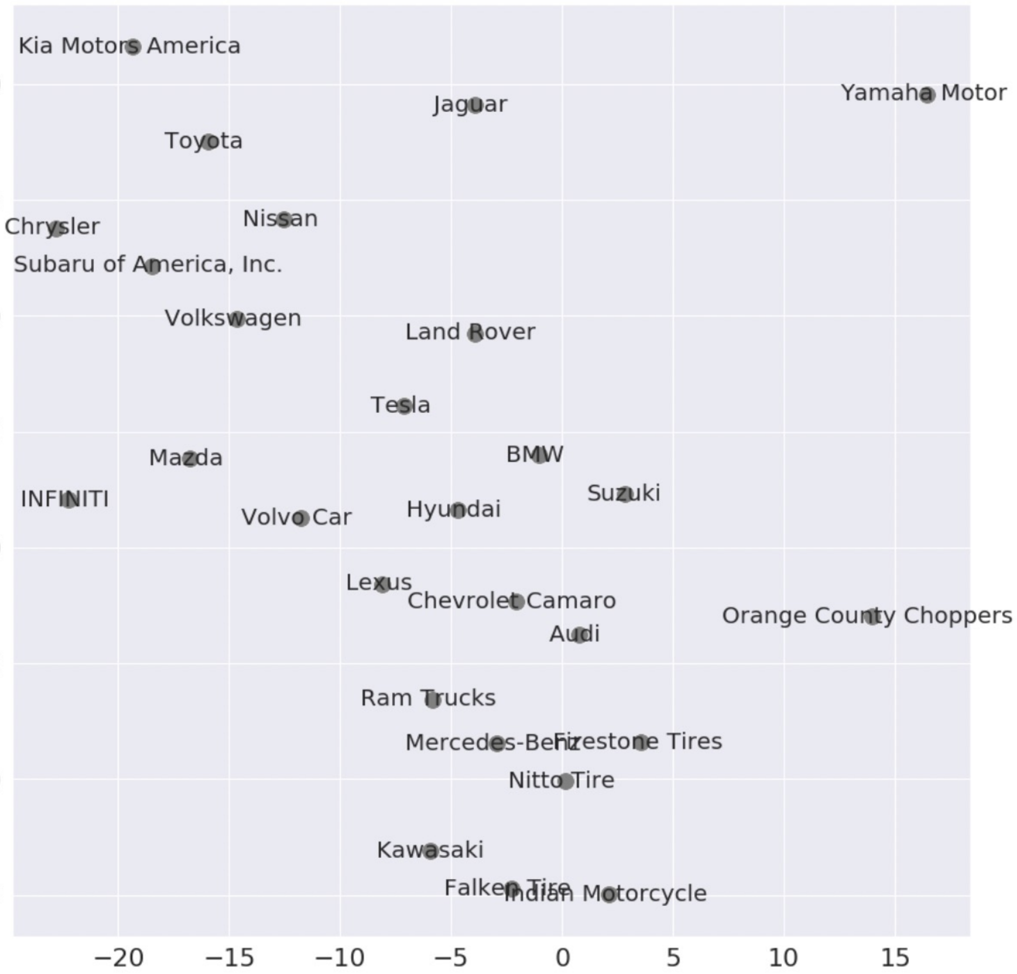
“The Luxury Travel Expert” is also close to “The Peninsula Beverly Hills,” a 5-star hotel

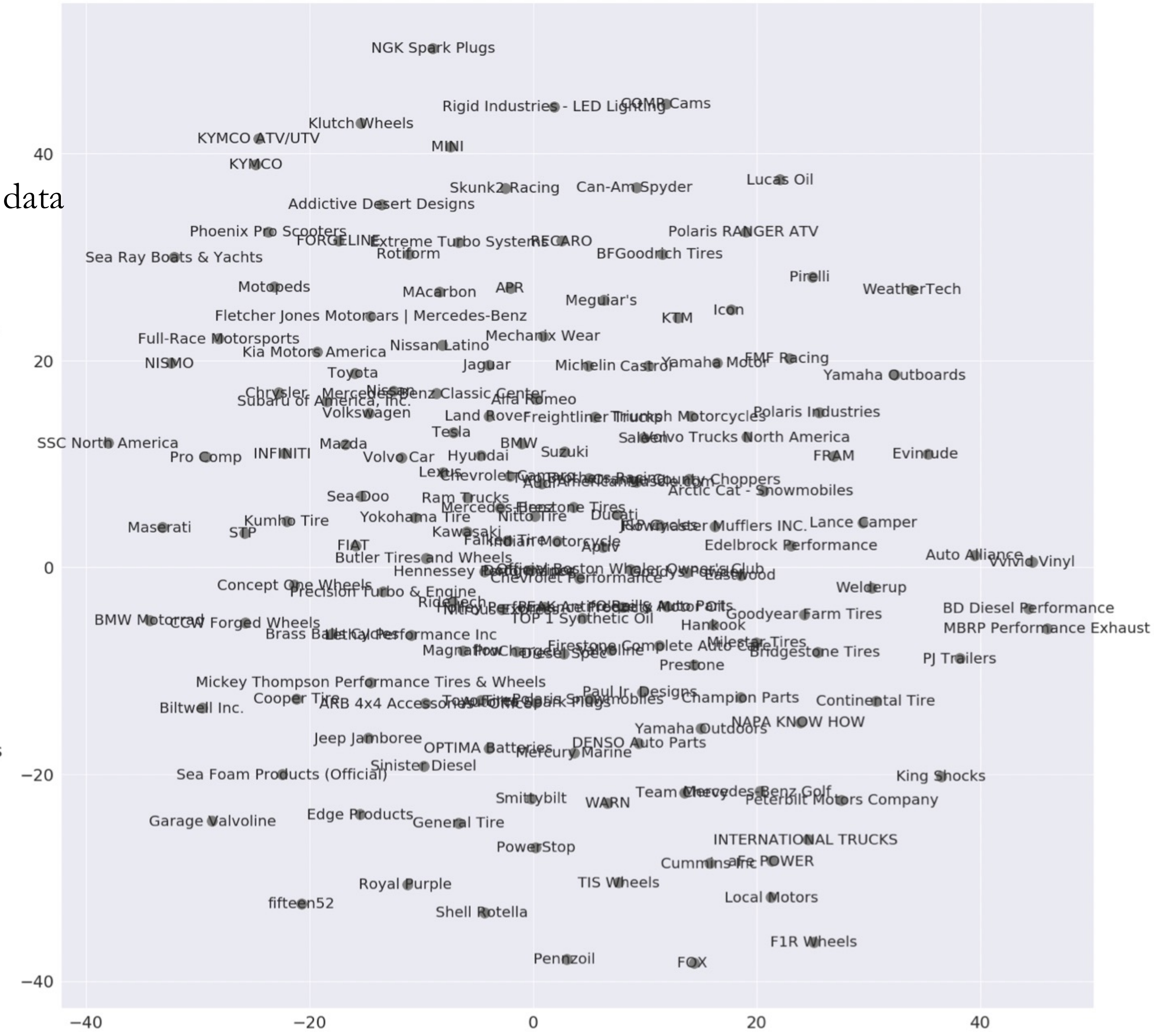
Visualization of market structure of 241 travel brands

Within-industry analysis

Visualization of market structure of 163 auto brands

Less clustered and more ambiguous compared to using all data





Within-industry analysis

"**FMF Racing**" - is a company that develops dirt bike exhausts for off-road or racing motocross riding

Top 10 proximal brands derived using engagement data from 'auto' brands only:

"Lucas Oil"
"KTM USA"
"Yamaha Motor"
"Arctic Cat"
"Two Brothers 22 Racing"
"Phoenix Pro Scooters"
"Auto Alliance"
"Valvoline USA"
"Lance Camper"
"Castrol"

"Lucas Oil," "Valvoline USA," and "Castrol" are global automotive oil brands

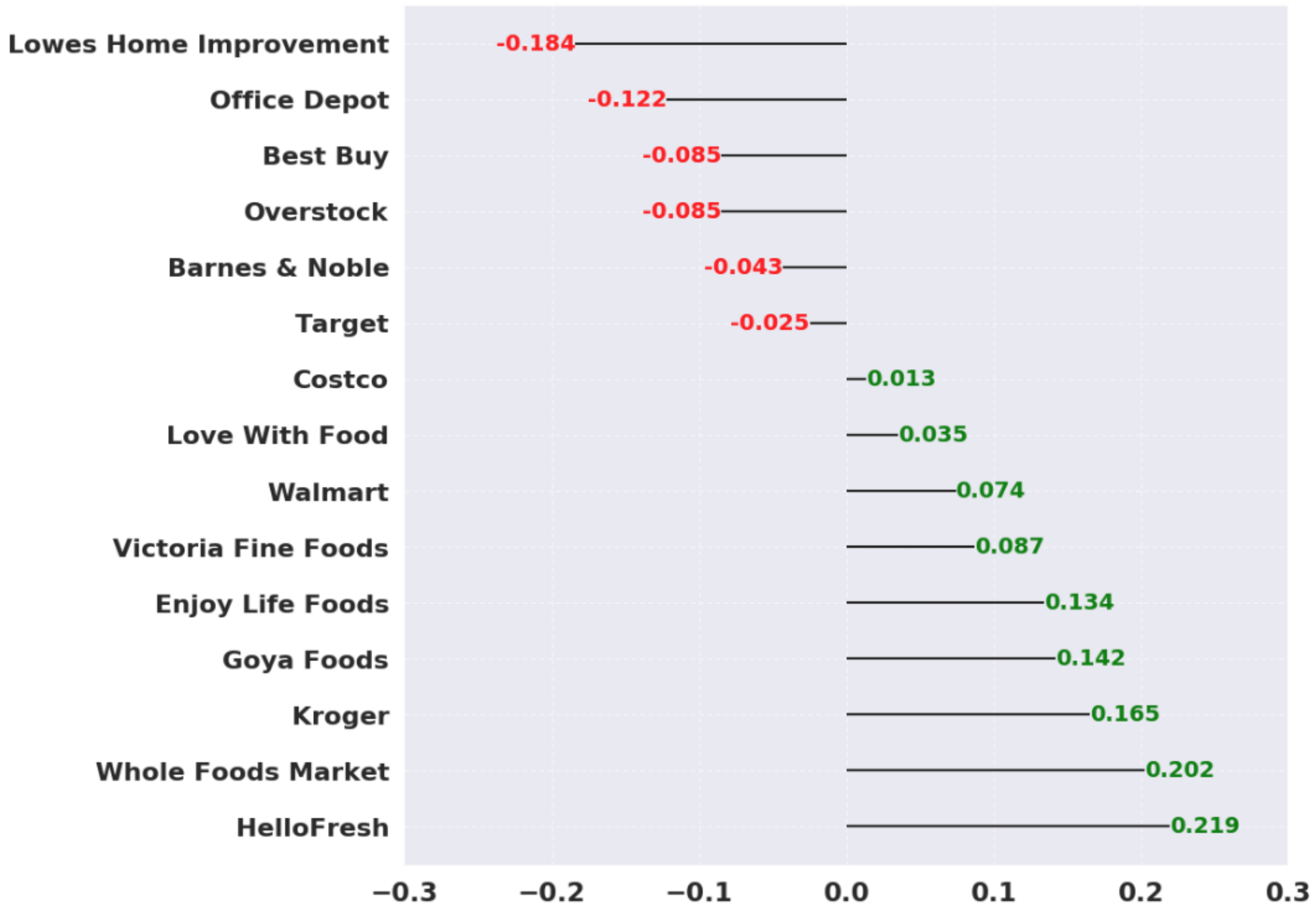
Top 10 proximal brands derived using engagement data from all brands:

"KTM USA"
"Polaris Snowmobiles"
"Fox Racing"
"Mickey Thompson Performance Tires & Wheels"
"Two Brothers Racing"
"King Shocks"
"Arctic Cat"
"Addictive Desert Designs"
"NISMO"
"Skunk2 Racing"
"MBRP performance exhaust"

All related to off-road motocross riding

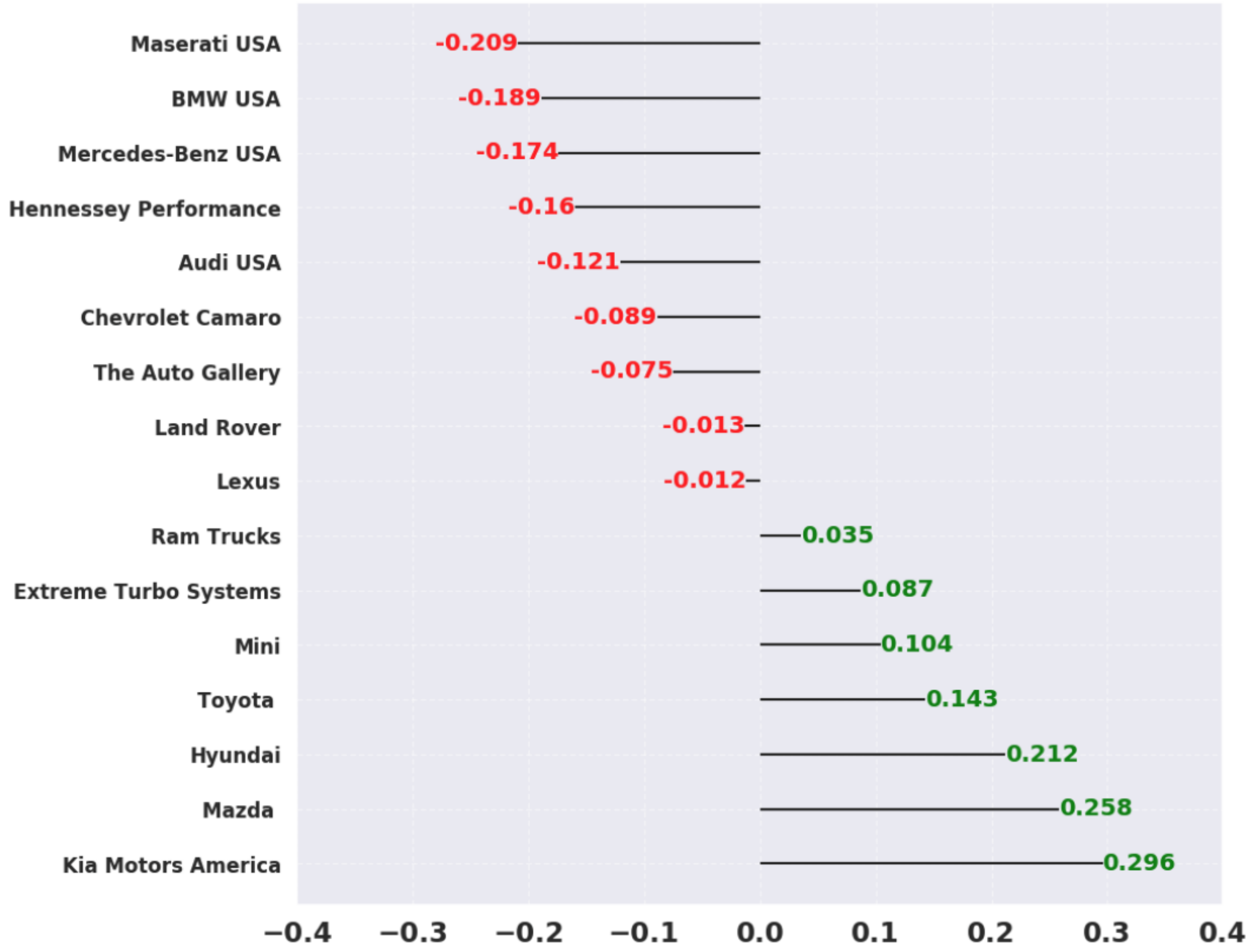
Case study

- Amazon acquires Whole Foods (August, 2017)



Case study

- Tesla delivers model 3 (July, 2017)



Conclusions

- Develop deep network representation learning on large-scale social media data for market structure discovery
- Add on to existing research on market structure discovery from a network analysis perspective
- Able to pin a large amount of brands on the market structure map to precisely visualize brand relationships
- Showcase how new technology can be used to better tackle a traditional marketing task

Conclusions

The research contributes to understanding the market boundaries and overlaps among different product categories

Dynamic analysis of changes in market structure and boundaries

Different implications of likes, comments and shares?

THANK YOU!