Image Network and Interest Group – A Heterogeneous Network Embedding Approach to Analyze Social Curation on Pinterest

Liye Ma

Baohong Sun

Kunpeng Zhang[1]

November 2019

[1] Liye Ma is an Associate Professor of Marketing at Robert H. Smith School of Business at the University of Maryland, College Park. Baohong Sun is the Distinguished Chair Professor of Marketing at Cheung Kong Graduate School of Business. Kunpeng Zhang is an Assistant Professor of Decision, Operations & Information Technologies at Robert H. Smith School of Business at the University of Maryland, College Park.

.

Image Network and Interest Group – A Heterogeneous Network Embedding Approach for Analyzing Social Curation on Pinterest

Abstract

Social curation platforms help consumers navigate through vast digital content to find what fits their interests. However, little is known about this important phenomenon. Using the popular image curation site Pinterest.com as empirical context, this research aims to understand: (i) how digital content is organized at social curation platforms; (ii) what users' curation activities reveal about consumer preferences, content characteristics, and brand positioning; and (iii) how we can predict users' curation actions.

We propose a novel approach which has two key components. First, we represent social curation using a heterogeneous information network. Images, users, and curation words are represented as nodes, while edges represent users' collection and annotation actions. Second, we leverage heterogeneous network embedding, a recently developed machine learning method, to map network nodes to lower-dimensional vectors, while preserving the network's structural and semantic information. We then analyze the embedding vectors for prediction and interpretation.

Analyzing a large furniture-related dataset from Pinterest, our proposed approach significantly outperforms prevailing benchmarks on predicting users' curation actions. Furthermore, embedding results reveal various user interest groups and image clusters, each with distinct characteristics. The match between users and images is stable out-of-sample. The analysis also generates insights on brand positions.

Keywords: social curation, information network, network embedding, machine learning, image processing, interest group, visual marketing, targeting

1. Introduction

The rapid accumulation of vast digital content on the Internet has led to the emergence of popular *social curation* websites such as Digg, Reddit, Scoop.it, and Pinterest. These social curation sites help consumers sift through the massive and unstructured content across the web, in text, image, audio, and video formats. Several *unique characteristics* of social curation sites make them an interesting and important phenomenon in industry. First, unlike social media sites such as Facebook and Twitter that focus on connections among people, a social curation site focuses on content, providing tools for users to discover, collect, organize, and annotate content such as news, articles, and photos. Second, rather than rely on domain experts to make curation decisions as is traditionally done, on a social curation website, content curation is crowdsourced, resulting from the collection and annotation actions of everyday users. [2] Third, the curation of one user is typically publicly available to other users. This collaborative environment makes it easy for everyone to discover the content that fits their interests. Finally, since millions of users individually collect and annotate content according to their own interests, a social curation site contains a vast amount of information on consumer preferences and content characteristics that is vital to marketers. Thus, social curation sites are an increasingly indispensable component of the Internet ecosystem, and an important space for knowledge discovery.

The most popular social curation site is Pinterest, a visual *folksonomy* platform that allows users to *pin* images into *boards* of related images that they maintain on the website and *annotate* according to their own preferences.[3] Users' pins and annotations of images collectively form the social curation process. Launched in 2010, Pinterest has quickly become a "catalog of ideas" to which users flock to discover, organize, and share images. Images are either collected from other websites by users, or uploaded to the site directly, sometimes by firms to advertise their products or services. Images are actively viewed and curated by users on the website, with some images curated thousands of times or more. Pinterest is a transparent site; all the information is viewable and collectible by other users and the general web-public. With over 175 billion pins and 250 million monthly active users, Pinterest is one

[2] Social curation differs from "traditional" curation in that the curation process involves more than one individual.
[3] The term "folksonomy" combines "folk" and "taxonomy," indicating that classification or annotation is provided by users.

of the most popular social media sites.[4] Since consumers frequently use the website to shape their ideas and guide product choices, Pinterest is a content repository with significant marketing implications. Analysis shows that two-thirds of the pins on the site represent products and brands, and consumer connections to brands on Pinterest are stronger than on traditional social network venues.[5]

An effective *visual marketing* strategy on social curation sites such as Pinterest has become an imperative for many firms. However, brands face a unique challenge in that while they can freely post images online, they have no control over their propagations. Curation actions (i.e., pins and annotations) are taken and controlled by consumers. Thus, understanding consumer preferences through their social curation actions is both important and challenging. However, academic research on social curation and its marketing implications is limited. Conceptually, users' curation actions are indicative of their preferences and interests, and of the match between their interests and the curated content. However, while social media research has either examined the creation and the effect of user generated content (e.g., Moe and Schweidel 2012, Stephen and Galak 2012, Tirunillai and Tellis 2012, Toubia and Stephen 2013, Tirunillai and Tellis 2014), or investigated the communication of consumers and firms on platforms of people network such as Facebook and Twitter (e.g., Ma et al. 2015), to date the rich information at social curation sites has largely been left unexplored. The importance of social curation platforms and the lack of in-depth understanding of them together motivate our research.

In this study, using Pinterest as the empirical context, we focus on analyzing users' social curation actions to generate insights into consumer preferences and conceptions to develop tools to extract and present such insights for marketers. We address the following research questions: First, what are the underlying structures of images and users as revealed by users' curation actions? When a user curates an image, the action reflects both the user's interest and the characteristics of the image, as well as the match between them. By analyzing all the curation actions, can we uncover groups of consumers who share common interests and clusters of images that have similar characteristics? Second,

[4] "*Which Social Media Platform Is The Most Popular In The US?*" URL: https://www.forbes.com/sites/kevinmurnane/2018/03/03/which-social-media-platform-is-the-most-popular-in-the-us/, accessed in December 2018.

[5] "Pinterest by the Numbers: Stats, Demographics & Fun Facts," URL: https://www.omnicoreagency.com/pinterest-statistics/, accessed in December 2018.

how can we describe users' interests and image characteristics? Curation goes beyond collecting images in that users also annotate the content or the collection action with words. What do annotations reveal about the interests of the users who curate the images and the characteristics of the images that are curated? Third, what can we learn about brands and their positioning as perceived by users based on their curation actions? Finally, and equally importantly, using the understanding of users' interests and image characteristics, how can we anticipate and predict which users are interested in particular products and what images they will curate? All these questions are important for generating consumer insights and sharpening brands' visual marketing strategies.

Addressing these substantive questions is uniquely challenging. Although the millions of images and users on Pinterest are interrelated in many aspects and dimensions, the interrelationship and structure are hidden and can only be understood through users' curation actions, i.e. *collection* and *annotation*. Such large-scale unstructured data are difficult to analyze using traditional choice modeling frameworks. Instead, we propose a novel approach in this study that includes two key components. First, we formulate a *heterogenous information network* representation for consumers' content curation activities. In the heterogeneous network, users, images, and annotation words are all represented as nodes, while users' curation actions (collection and annotation) are represented as edges that connect the nodes. Since each curation action reflects a consumer's interest in an image, this heterogeneous network presents a global view of social curation, and encodes rich information about consumer interests, image characteristics, and the alignment among them. Second, we analyze the heterogeneous network using a recently developed method in the field of machine learning: heterogeneous *network embedding* (Dong et al. 2017). Using network embedding, each node in the network is mapped to a point (i.e., vector) in a vector space, such that the structural and semantic information of the original network is preserved in terms of the distances among the vectors in the vector space. With the extracted structural and semantic information in these mapped vectors, we then apply existing methods that take vectors as inputs to further extract insights on consumers, content, and brands.

Our proposed approach has several attractive features. To begin, the heterogeneous network representation provides a global view of users' organization of images and the related context. By putting user, image, and annotation into one multimodal network, we are able to incorporate user interactions with images and curation words in an integrated framework with little loss of information. Next, by using the network embedding method, we can effectively capture the structural and semantic relationship exhibited from such a large-scale network with multiple entity types. This allows us to establish associations between images, users, and annotations that are linked both directly and indirectly, since both local and global information of the network is preserved. Furthermore, the structural and semantic information is captured using vectors in a vector space, making it straightforward to perform additional analyses to draw insights about consumer interest groups, image clusters, and brand positioning. Finally, the extracted structural and semantic information can be directly utilized in prediction models, which can help brands or social curation sites recommend matching content to interested users.

We apply the proposed approach to a large dataset that contains over one-thousand furniture-related images and their curation actions by more than one-hundred-thousand users over three months in 2017. The proposed approach handles high volumes of data with ease. We identify 15 user interest groups and 8 image clusters each with distinct characteristics. By associating users and images with annotation words, we further describe the consumer preferences and image characteristics of each identified group. We also analyze the connections between the user and image groups, and show that the extracted relationship is highly stable. Furthermore, by analyzing branded images, we create a brand positioning map showing the central and peripheral positions of different brands, and analyze the extent of brand cohesion. Equally importantly, our proposed approach performs significantly better in predicting users' future content curation than benchmark approaches that are commonly used in the industry, such as collaborative filtering, matrix factorization, social network-based methods, and community detection-based methods.

Being the first to investigate social curation and to apply network embedding to marketing problems, our research contributes to the literature in two important ways: First, methodologically, our study is the first to represent social curation using a heterogeneous information network, and to introduce the network embedding method to the marketing literature. The heterogeneous network representation effectively presents a global and holistic view of the platform-level information in an integrated setting. Furthermore, network embedding is a powerful approach that projects nodes in a large-scale network to vectors in a low-dimensional space, while preserving the semantic and structural information of the network. We show that our proposed approach generates easy-to-analyze embedding vectors, draws important insights from them, and achieves significantly better predictive performance than popular extant benchmark methods. Our approach is thus effective in both capturing information from big data and generating insights from it. Furthermore, this proposed approach also has broader potential applications.

Second, substantively, using Pinterest as the empirical context, we shed light on the social curation process. We extract meaningful user and image groups, establish their interrelationships, and use annotations to describe the nature of user interests and image characteristics. Putting brands on the map further assists firms in engaging with consumers on social curation sites. These insights add to our knowledge of this important yet under-studied phenomenon. We also provide a rich set of tools for prediction, clustering, text analysis, and brand positioning and cohesion analysis that can directly assist marketing managers in their analytics efforts, further enhancing the substantive contribution of this research.

2. Literature

Our study is related to several streams of literature including networks, network embedding, social curation, and image processing.

2.1 Social media and network

A rich marketing literature exists on social media and social networks. Studies have investigated various aspects of word-of-mouth (WOM) on the Internet. Two large streams of work in this literature show the important effects of WOM (e.g., Godes and Mayzlin 2004, Chevalier and Mayzlin 2006, Godes and Mayzlin 2009, Trusov et al. 2009, Chintagunta et al. 2010, Stephen and Galak 2012, Tirunillai and Tellis 2012, Tirunillai and Tellis 2014), and shed light on the underlying factors that motivate the creation of WOM (e.g., Anderson 1998, Hennig-Thurau et al. 2004, Schlosser 2005, Berger and Schwartz 2011, Berger and Milkman 2012, Toubia and Stephen 2013), respectively. Other studies have investigated the dynamics of WOM, online sentiment analysis, and complaint management in social media (Li and Hitt 2008, Moe and Trusov 2011, Godes and Silva 2012, Moe and Schweidel 2012, Schweidel and Moe 2014, Ma et al. 2015).

A number of studies have also investigated online networks. Katona and Sarvary (2008) analyzed the hyperlinks among websites. Stephen and Toubia (2010) found that sellers in an online social-commerce marketplace derive significant benefit from connection with peers, and this benefit primarily comes from the accessibility enhancement of the network. Studies have also analyzed the network structure and its modifying effect on consumer decisions. Zhang and Godes (2017) showed that a bidirectional relationship represents a stronger tie than the unidirectional link. Lu et al. (2017) developed a dynamic game model to endogenize the formation of social network in an online crowdsourced customer support platform. Kumar and Sudhir (2019) investigated WOM seeding strategies in social networks, and showed that strategies leveraging the friendship paradox are more effective than opinion-leader seeding.

Our research is related to this literature, but differs from it in two important ways. First, while existing studies have looked at the creation or the effect of user-generated content, our study focuses on a different aspect, the *curation* of content. The curation actions taken by consumers contain rich information on consumer preferences and content characteristics, as well as the match between them. Our study seeks to analyze this rich information, which has not been explored in the literature. Second, while existing network studies have focused on *social networks of people*, we construct a heterogeneous *information network* in which users, images, and annotations are all treated as nodes, and edges represent the curation actions. Instead of just encoding friendship connections as in a typical social network, the heterogeneous information network proposed in our study captures a more general view of the multiple types of entities resulting from curation actions, thus providing insights to guide brands' visual marketing efforts.

2.2 Pinterest and social curation

Despite the relatively recent introduction, Pinterest has begun to attract research interest among marketing scholars. The limited extant literature has focused on examining the overall motivations for using Pinterest. Hall and Zarro (2012) were the first to recognize Pinterest as a social curation and information literacy tool. They concluded that Pinterest is a sharing and curating platform, which is beneficial for information use, reuse, and creation on the social web. Miller et al. (2015) provided evidence that Pinterest users tend to view Pinterest as a place to discover, collect, and share with others. Zhong et al. (2015) sought to understand Pinterest as distributed human computation that categorizes images from around the Internet, and presented a coarse-grained taxonomy of 32 image categories. Our study seeks to substantially advance this nascent literature on social curation, by revealing insights about consumers' preferences and content characteristics based on users' collection and annotation of images.

2.3 Network embedding

Large networks with millions or billions of nodes are becoming commonplace. However, one of the challenges when dealing with such large networks is to find effective approaches to represent networks concisely and efficiently, so that advanced analytic tasks, such as pattern discovery, prediction, and inference can be subsequently conducted. Working with the network connection structure usually involves iterative or combinatorial operations, leading to high computational complexity. Since the nodes in a network are related to each other as encoded by the edges, most analytical methods, which assume that data samples are represented by independent vectors in a vector space, cannot be directly applied.

To overcome these challenges, machine learning researchers in computer science have developed an efficient method, *network embedding*, to learn low-dimensional vector representations for network nodes. Network embedding (Zhu et al. 2007, Tong et al. 2008) aims to project the structural proximities of all nodes in a network into a continuous low-dimensional vector representation. The learned embedding representation paves the way for numerous applications such as node classification, link prediction, and network visualization, enabling both accurate prediction and insightful knowledge extraction.

The pioneers of network embedding date back to the 2000s when many graph embedding algorithms were proposed, e.g., an unsupervised locally linear embedding introduced in Roweis and Saul (2000). These methods first build an affinity matrix that preserves the local geometry structure of the data manifold, and then embed the data into a low-dimensional representation. Motivated by the graph embedding techniques, Chen et al. (2007) proposed one of the first network embedding algorithms for directed networks. They use random walk to measure the proximity structure of the directed network. Recently, network embedding techniques have received a surge of research interest. Among them, Deepwalk (Perozzi et al. 2014) generalized word embedding and employed a truncated random walk to learn latent representations of a network. Node2vec (Grover and Leskovec 2016) further extended Deepwalk by adding flexibility in exploring node neighborhoods. LINE (Tang et al. 2015) carefully designed and optimized objective function that preserves the first-order and the second-order proximities to learn network representations. GraRep (Cao et al. 2015) improved LINE by taking high-order information into account. In addition, struc2vec (Leonardo, et al 2017) proposed a hierarchy to measure node similarity at different scales, and constructed a multilayer graph to encode structural similarities for nodes. The literature has grown rapidly since then, with extensions to analyze complicated networks such as attributed networks (Huang et al. 2017, Li et al. 2017), dynamic networks (Li et al. 2017), heterogeneous networks (Chang et al. 2015, Dong et al. 2017), and attributed signed networks (Wang et al. 2017). More recently, deep-learning based approaches have also been proposed (Wang et al. 2016, Yang et al. 2016). For example, Gao and Huang (2018) proposed a deep attributed network embedding approach to capture high nonlinearity and preserve various proximities in both topological structure and node attributes. Generalizing well-established neural network models such as CNNs or RNNs to deal with arbitrarily structured graphs is challenging. Thus, recent work focuses on alleviating this gap by developing spectral approaches (e.g., graph convolutional networks) where researchers achieved convincing results on a number of benchmark graph datasets (Kipf and Welling, 2017, Defferrard, et al. 2016). Specifically, Dong et al. (2017) proposed metapath2vec, which builds upon earlier skip-gram models of homogeneous network embedding to explicitly account for node types in generating embedding vectors, (i.e., the method generates embeddings for heterogeneous networks). This is the method adopted in our study.

2.4 Image processing

It is well-known that analyzing image data is challenging. However, recently, the field of image processing has made significant advances using deep neural networks (e.g., Google Cloud Vision API), and image analysis will likely become a new frontier in marketing research. Studies have begun to adopt advanced machine learning approaches to process images. Liu et al. (2017) proposed a "visual listening in" approach to measure how brands are portrayed on the social media site Instagram, by mining visual content posted by users. Two supervised machine learning methods were used to measure brand attributes (e.g., glamorous, rugged, healthy, fun) from images, and the authors found key differences between how consumers and firms portray the brands on visual social media. Zhang et al. (2017) analyzed the effect of images on property demand at Airbnb. Using a deep-learning based algorithm to classify images as high quality or low quality, they showed that having verified photos with high image quality increases demand. Li et al. (2019) also used deep learning methods to extract measures from videos. They showed the measures can predict project funding outcomes at a crowdfunding site. While existing literature on imaging processing mainly analyzes image content based on the pixels, our study takes a different approach by analyzing consumers' curation actions. We also use image processing tools to extract content characteristics, similar to what has been done in this stream of research, although in our study the extracted characteristics are used in a post hoc manner to draw descriptive insights, while our primary approach centers on representing and analyzing social curation actions.

3. Empirical Context and Data

3.1 Pinterest

Pinterest is an open website where users can discover, collect, organize, and annotate images and other media content (e.g., videos). On the website, both an image and the action of collecting it are referred to as *pin*. Users either *pin* images from other websites to the Pinterest website, or upload the pins directly to the site. The images are organized into user-defined folders, or *pin boards*. Images on Pinterest span a wide range of categories: food, do-it-yourself (DIY) crafts, home decoration, fashion, health, fitness, among others. While pinning, users often add textual descriptions, based on which *tag words* are associated with images on the website. Figure 1 shows an example of a user's boards and an individual board. Pinterest is a transparent site, and all the information is viewable and searchable by other users and the general web-public. The same image can be pinned by thousands of users or more.[6] The name of each pin board to which an image is pinned provides information about the image. For each image, the website displays the number of pin actions, number of likes, and all comments. For each user, the website displays the username, a profile, number of boards, number of pins, and number of likes.

Images in the same pin board typically share a common theme, which may be reflected in the *board name*. The theme may be task-oriented, such as "porch design," object-oriented, such as "living room furniture," topic-oriented, such as "travel," or in any way the user prefers. *Annotations* in the form of tag words and board names provide meaningful information about users' interests. Through the social curation process, additional consumer annotations are added to the images, revealing both the interest of the consumer who provides the annotation and the characteristics of the image.

[Insert Figure 1 About Here]

Industry statistics show strong marketing implications of Pinterest. For example, 93% of users use Pinterest to plan purchases, 72% use Pinterest to decide what to buy offline, and 87% of users have purchased something because of Pinterest.[7] Pursuing visual marketing strategies, brands have flocked to Pinterest to create virtual storefront, posting images of new collections, curating seasonal looks, and inspiring consumers. In addition, brands frequently add a "Pin-it" button on promotional images on their own websites to encourage users to collect their images to Pinterest. However, while they can supply the images, brands have no control over their disseminations, which are completely driven by consumers' curation actions. Thus, a key challenge for firms is how to identify interested consumers and present them with images that match their interests and preferences.

Unlike Facebook and Twitter that rely on explicit people networks for information diffusion, curation actions on Pinterest do not rely on the existence of a social network. Pinterest does allow users

---

[6] On the website, the first time an image is collected to the website it is called *pinned*, and subsequent collection actions by other users are called *repins*. In this study, we do not make this technical distinction, as they both represent curation actions in the same manner. Due to technical limitations, the initial pin action of an image is not tracked in our dataset. This is considered a minor issue, since repin actions are far more frequent than the initial pin actions.

[7] "Pinterest by the Numbers: Stats, Demographics & Fun Facts," URL: https://www.omnicoreagency.com/pinterest-statistics/, accessed in December 2018.

to follow each other, thus creating a following-follower network relationship, much like the one on Twitter. However, this connection among people on Pinterest is secondary compared with Pinterest's main focus: users' collection and annotation of content.

3.2 Data

For this study, we obtained a dataset of furniture-related images and the corresponding curation actions from the Pinterest website as follows. For each day from 3/1/2017 to 5/31/2017, we searched the keyword "furniture" on the site. From the search results, we recorded the images that first appeared on the site on that day. Through this approach, we collected a comprehensive set of furniture-related images first posted on the site within the designated three months. The raw image file together with a textual description and a set of *tag words,* where available, were retrieved for each image.[8] We then downloaded the entire pin history of all the images in the dataset for the same three-month period. Each pin record contained the ID of the user who performed the pin action, the date the pin action took place, and the name of the pin board to which the image was pinned. We also downloaded the characteristics of all the users who pinned the images in the dataset, including the initial registration date, the number of users they followed, the number of their followers, the total number of pins, and the total number of boards. These data were downloaded using the API[9] provided by Pinterest.com and from scraping the webpages directly on the website.

We used the first 8 weeks covered in the dataset for the network embedding analysis (up to and including 4/25/2017), and the remaining time period for holdout evaluations. The initial dataset contains a total of 2,004 images each with at least one pin action available. Among the 2,004 images, 1,089 were first posted to the site in the first 8 weeks of the dataset. The initial dataset contains 275,396 users who have pinned some of those images in the three months covered by the dataset. Among them, 108,123 users first pinned some of those images in the first 8 weeks. Our final training dataset (i.e., the dataset used for the network embedding analysis) consists of these 1,089 images and 108,123 users.

[8] Tag words are associated with image files on the website, presumably based on user descriptions and comments. The exact method used by the website to create such associations is unknown. In this study, we assume that tag words are distilled user annotations.
[9] https://developers.pinterest.com/docs/getting-started/introduction/?

[Insert Table 1 and Figure 2 About Here]

The images were actively pinned by users. The average number of pin actions per image is 112, with a standard deviation of 203 and maximum of 1,311. The histogram of the number of pin actions by image is shown in Figure 2. As is common in such data, the distribution is skewed, with a majority of the images receiving fewer than 100 pins each, while a small set of images were pinned more than 1,000 times over the three-month period. The descriptive statistics of users are reported in Table 1. Users in our sample are, in general, active on the website, although their levels of activity vary significantly across users. The number of pin boards each user has ranges from 1 to 3,038, averaging 52.73. On average, each user pinned a total of 9,748 images (of all images, not just those in our dataset). These statistics suggest a generally high level of activeness on the website. Furthermore, there is a well-connected social network on Pinterest, and the users in our dataset on average follow 357 other users, and have 1,166 followers.[10] The average user tenure (number of years the user has been on the site at the time of data collection) is 3.52 years. As Pinterest was only created in 2010, this suggests that our dataset has a balanced coverage of early and late adopters.

Tag words and board names are generated through users' social curation process. They typically describe the images' purposes, applicable settings, styles, qualities, as well as other characteristics. There are a total of 64 unique tag words (after standard text preprocessing, including stopword removal and stemming) for the images in our training dataset. For the board names, we first extracted all unique words used by the users, and then included the 100 most frequent words (using the same preprocessing as tag words) in our analysis. The tag words and the frequent board names are referred to as *annotation words* in our analysis. The top 20 tag words, the top 20 board names, and their respective number of associated images are available in the Online Appendix (Table TA.3.1).

Figure 3 shows a few examples of the images in our dataset. While the images are all related to furniture, they differ in properties such as size, shape, color, and hue. For each image, we used the

[10] Note that while the numbers of follower and following are known for each user, the actual pairwise connections between individual users are not available, due to the large volume. We were able to obtain the social network connections for only a small subset of users, which we use for benchmark comparison. We show in section 5 that the information network proposed in this study outperforms the benchmark method which relies on the social network among users.

image processing tools OpenCV and Google Vision API to generate a set of image characteristics such as pixels, length, width, and brightness. We also used the crowdsourcing platform Amazon Mechanical Turk (AMT) to identify the brand each image is associated with. For each image, we asked 10 different workers to identify the brand (10 brands plus other). The final brand was summarized based on the majority vote. A total of 715 images in the training sample and 527 images in the holdout sample were each associated with one of ten brands through this approach. The brands and the number of images are reported in Table 2. IKEA and Restoration Hardware are the two brands with the most images, with more than one-hundred each in the training set.

[Insert Table 2 and Figure 3 About Here]

4 Representing and Analyzing Social Curation

We now discuss the representation of social curation using a heterogeneous network and the analysis using network embedding. Although our discussion is based on the context of an image curation platform like Pinterest, our proposed approach is potentially generalizable to other settings.

4.1 Conceptual background of heterogeneity networks and network embedding

Conceptually, a user curates an image on Pinterest when the image matches her interest in certain aspects. A curation action thus indicates the match between the user's preference and the image's characteristics, and annotation words provide additional description of this match. Understanding user preferences and image characteristics from the curation process is crucial to brand firms as well as the content platform. However, the large-scale unstructured data present a formidable challenge.

To overcome this challenge, and to derive a global and holistic view the social curation on Pinterest, we propose a novel analysis approach, which consists of *two key components*. First, we represent the entities on the website, namely images, users, and annotation words, as well as the curation actions (i.e., collection and annotation) that connect the entities, using a *heterogeneous information network*. Second, we use the *network embedding* method in machine learning to map the nodes in this heterogeneous network to vectors in a lower-dimensional vector space, while preserving the original network's structural and semantic information in the distances between the embedding vectors. We then analyze the embedding vectors for prediction and interpretation.

We propose this new approach, instead of using a conventional choice modeling method, for several important reasons. First, choice models take the perspective of decision-making individuals, and are not aimed at deriving a global view of all the entities and their relationships, which is our research objective here. Second, unlike a regular product with clearly-coded features, a digital image is a collection of pixels that are unstructured and without well-defined characteristics. While advancements in image processing methods have made meaningful feature extraction possible, the extracted features are usually technical in nature, (e.g., brightness and hue) rather than contain consumer preference information. Thus, finding meaningful independent variables for the choice model is difficult. Third, while curation words can be treated as independent variables, such usage is questionable since curation words reflect a user's *subjective* perception of the image characteristics. For example, some users may see "outdoor" in an image, while others may see "cool" or "contemporary." Finally, the large number of images at Pinterest makes it too computationally demanding for even minimally specified choice models such as Probit or Logit. To use choice models, strong simplifying assumptions have to be made, which will limit the extent of insights that can be generated.

Our proposed approach addresses these challenges, making it a better fit for the context of this study, and a useful alternative in general to complement the traditional choice models. In the first key component of the proposed approach, we construct a heterogeneous network, in which *users*, *images*, and *annotation words* are all represented as *nodes*, while users' *curation actions* are represented as *edges* that connect the corresponding nodes. This heterogeneous network elegantly incorporates users' curation of images with little loss of information, while providing a global view of all the entities and their interrelationships as collectively defined by users. Users who are close to one another in this network likely share similar interests, as they curate similar images and provide similar annotations. Thus, interest groups can be discerned from clusters of users that are closely connected in the network.[11] Similarly, common characteristics of images can also be discovered from image clusters.

The middle pane of Figure 4 illustrates the heterogeneous network representation, where U, I, and A represent user, image, and annotation word, respectively. There are five users who curate four

[11] This is different from a people network such as Facebook, where connections encode relationships among users and clusters are likely to represent social communities.

images: users U1, U2, and U3 curate image I1; U1 and U4 curate image I2; U4 and U5 curate image I3; U5 curates image I4. Furthermore, while curating image I1, U1 also uses annotation word A1. Thus node A1 is connected to U1 and I1. The same is true for U4 who uses the annotation word A2 while curating images I2 and I3. Intuitively, we can see that users U1, U2, and U3 have similar preferences that match the characteristics of image I1, which is reflected in the annotation word A1. Furthermore, I3 likely has certain characteristics that U4 and U5 both prefer. Finally, image I2 would have some characteristics that appeal to some users in both groups. All these insights are elegantly encoded into the heterogeneous network.

[Insert Figure 4 About Here]

While this network effectively incorporates the social curation information, obtaining insights from the network is not straightforward. A real-world network can contain millions of nodes, which precludes visual analysis. More importantly, a network is not an object in a metric space, and standard classification or regression methods are not directly applicable. A naïve conversion of network to metric space is possible. For example, one can represent each user using a separate dimension, and represent images and annotations as vectors in this space, with each dimension enocding whether the image/annotation is connected to the corresponding user. Alternatively, each dimension can represent an image, and users and annotations are represented as vectors in this space. In both cases, however, the vectors have high dimensionality, equaling the number of users and images, respectively. While simple operations are feasible, this high dimensionality precludes more in-depth analysis. Alternatively, extant literature often analyzes networks in an ego-centric way using node-level variables, e.g., the degree centrality. More detailed information at the edge- or node-pair level may also be used, e.g., the connection strength between two nodes. However, this approach only leverages the *local* information of direct or indirect neighbors, while leaving out the large amount of information contained in the global network structure. As another alternative, methods such as community detection account for the structure of the network, but they typically involve a degree of arbitrariness, e.g., to define the in-group cohesion to be maximized, while achieving only limited extraction of the structural information.

The heterogeneous network embedding method in our proposed approach offers a more desirable alternative to overcome these limitations. Using network embedding, each node in the network is mapped to a vector in a vector space, such that the *closeness* between the nodes in the network is preserved through the *distance* between the embedding vectors in the embedding space. The embedding vectors have much lower dimensions than the original network, and existing methods that take vectors as inputs, e.g., clustering or classification models, can be readily applied to further extract insights from them. More importantly, network embedding preserves the structural and semantic information of the network in the embedding vectors. For example, nodes that are directly connected or share many neighbors in the network would see their embedding vectors have short distance in the embedding space.

The right pane of Figure 4 illustrates the intuitive appeal of network embedding. It maps the heterogeneous network in the middle pane to a 3-dimensional vector space. Reflecting the similarity in user interest and image characteristics, the embedding vectors for A1, I1, U1, U2, and U3 are close to each other, so are the vectors for U4, U5, I3, and I4, while that for I2 lies somewhere in between.[12] Furthermore, this vector representation enables more sophisticated and in-depth analyses. For example, a clustering analysis may show that there are two groups of distinct preferences among users and images. In large-scale, real-world networks such as the one constructed in our study, visual inspections of the network are no longer possible. In contrast, quantitative analysis of the embedding vectors is still effective at extracting insights.

In summary, the heterogeneous network representation and network embedding together capture and present key information about consumer preferences and image characteristics in a readily analyzable format. This is the key advantage of our proposed approach. We now proceed to the technical details.

4.2 Heterogeneous network representation

Formally, let there be J *images*, each denoted as $I_j, j = 1, \ldots, J$. Let there be N *users* registered on the website, each denoted as $U_i, i = 1, \ldots, N$. There are K *annotation words*, each denoted as $A_k, k = 1, \ldots, K$.

[12] For clarity in visualization, the labels are not displayed in the vector space plot.

We represent the entities and social curation actions as a heterogeneous network $G = (V, E, T)$, where V represents the vertices, or nodes, E represents the edges (links) between nodes, and T represents the types of nodes and links – each node v and edge e is mapped to its type: $\varphi^v: V \rightarrow T_V; \varphi^e: E \rightarrow T_E$. In our setting, there are three types of nodes: $T_V = \{U, I, A\}$, representing *user*, *image*, and *annotation*, respectively. User and image nodes are self-explanatory, while annotation nodes include all the tag words and the 100 most frequently used words in pin board names.

The type of an edge is determined by the types of the nodes it connects. For ease of exposition, we use V_i^u to represent a user node, V_j^i to represent an image node, and V_k^a to represent an annotation node. The following types of edges can exist: 1) that between a user V_i^u and an image V_j^i: $E_{ij}^{ui} = 1$ if the user V_i^u pinned the image V_j^i, and $E_{ij}^{ui} = 0$ if not (i.e., no link between the two nodes); 2) that between a user V_i^u and an annotation V_k^a: $E_{ik}^{ua} = 1$ if the annotation word V_k^a appears in the name of a pin board of the user V_i^u, and $E_{ik}^{ua} = 0$ otherwise; 3) that between an image V_j^i and an annotation V_k^a: $E_{jk}^{ia} = 1$ if the annotation word V_k^a is a tag word associated with the image or is in the name of a user's pin board which contains the image V_j^i; $E_{jk}^{ia} = 0$ otherwise. Annotation nodes are thus connected to both user and image nodes in the network.

The network defined this way contains all the available information about consumers' image curations.[13] The links to annotation nodes encode information about user preferences and image characteristics. For example, if a user is connected to the annotation "garden" by using the word in naming a pin board, it would suggest that the user is interested in garden design. Similarly, an image connecting to "garden" would indicate that the content of the image is related to garden design. Aside from interpretable annotation words, the connection structures themselves are informative. For example, the simple existence of a link between a user and an image indicates that the user's preference matches the image's characteristics. Indirect connections also indicate relations. For example, two

[13] Note that although the network in our study contains edges only between nodes of different types, it can easily be extended to include other types of edges. For example, should a social network structure exist among users, another link type can be introduced to represent such social connections. As another example, if relationships are deemed to exist between annotations such as a certain pair of words that have a general-specific relationship, additional link types between annotation nodes can also be introduced accordingly. The proposed heterogeneous network representation is both flexible and generalizable.

images are more closely related when they are more frequently curated by the same users and share more common annotations, i.e., when they share more neighbors in the network. We next discuss the use of network embedding to extract insights from this heterogeneous network.

4.3 Network embedding

Network embedding is a powerful method recently developed in machine learning. Mathematically, embedding is an injective and structure-preserving map. As such, network embedding is a mapping $f: V \rightarrow R^K$, where $K \ll |V|$, such that structural and semantic information of the original network is preserved. The approach to preserve information varies across methods. Tang et al. (2015) proposed the *LINE* method, one of the first network embedding methods applicable to large-scale networks, which treats a network as homogeneous, where only the connection structure is considered while node and edge types are ignored. The embedding preserves first-order and second-order proximity. The *first-order* proximity refers to direct connections between nodes. The probability of two nodes being connected is specified as[14]

(1) $$p_1(v_i, v_j) = \frac{1}{1+\exp(-\vec{u}_i^T \cdot \vec{u}_j)}$$

In Equation 1, v_i and v_j are two nodes in the network. In our setting, they could be the user, image, or annotation nodes. $\vec{u}_i$ and $\vec{u}_j$ are the K-dimensional embedding vectors corresponding to the nodes. Intuitively, two connected nodes should have embedding vectors with high cosine similarity, which is reflected in their dot product. The vectors are chosen to optimize the following objective function:

(2) $$O_1 = -\sum_{(i,j)\in E} w_{ij} \log(p_1(v_i, v_j))$$

In Equation 2, O_1 represents the objective function to be minimized to preserve the first-order proximity, and w_{ij} is the weight of the edge in the original network (0 or 1 in our setting).

[14] Since network embedding is a new approach to the marketing literature, we reproduce here certain key equations in the original articles for ease of reference. Additional technical details can be found in the original articles.

Second-order proximity refers to the similarity of neighbors between two nodes, for which the probability of context v_j generated by v_i is defined as

(3) $$p_2(v_j|v_i) = \frac{\exp(-\vec{u'}_j^T \cdot \vec{u}_i)}{\sum_{k=1}^{|V|} \exp(-\vec{u'}_k^T \cdot \vec{u}_i)}$$

In Equation 3, $\vec{u}_i$ is the embedding vector corresponding to node v_i, while $\vec{u}'_j$ is the representation of v_j when it is treated as a "context" to other nodes. The intuition is that if node v_i and another node, say v_k, have similar neighborhoods in that they both are connected to v_j, then both $\vec{u}_i$ and $\vec{u}_k$ are likely to be close to the context $\vec{u}'_j$, and in turn be close to each other. By capturing the second-order proximity this way, *LINE* goes beyond just relying on direct local connections. The vectors are chosen to optimize the following objective function:

(4) $$O_2 = -\sum_{(i,j)\in E} w_{ij} \log(p_2(v_j|v_i))$$

To preserve both the first-order and second-order proximities, *LINE* generates a vector for each node to optimize (2) and another vector for each node to optimize (4), and then concatenates the two vectors as the final embedding vector.

The *LINE* method performs embedding on a homogeneous network without consideration of node types. The network constructed in our study, however, is a heterogeneous network containing three types of nodes (user, image, and annotation), and this semantic distinction is important. For such a network, the *heterogeneous network embedding method* developed in Dong et al. (2017), is a better fit than *LINE*. Dong et al. (2017) extend earlier skip-gram models of homogeneous network embedding (Perozzi, et al. 2014, Grover and Leskovec 2016), which closely resemble the *word2vec* method (Mikolov et al. 2013) that precedes them. Following the skip-gram model, embedding is performed by taking repeated random walks along the network paths, and the embedding vectors are generated by maximizing the probability of having the heterogeneous context $N_t(v)$ for each node v:

(5) $$argmax_\theta \sum_{v\in V} \sum_{t\in T_V} \sum_{c_t\in N_t(v)} \log p(c_t|v;\theta)$$

In Equation 5, $N_t(v)$ denotes node v's *neighborhood* with the t-th type of nodes, and θ generically represents the parameter, i.e., the embedding vectors. The neighborhood is defined through random walks which is explained below. The probability in the equation is specified using the softmax function:

$$(6) \quad p(c_t|v;\theta) = \frac{\exp(X_{c_t} \cdot X_v)}{\sum_{u \in V} \exp(X_u \cdot X_v)}$$

In Equation 6, X_v is the K-dimensional embedding vector for node v. Intuitively, if one node is in another's neighborhood, equations 5 and 6 would seek to increase the cosine similarity between the two corresponding embedding vectors, i.e., reducing the distance between them.

To effectively preserve the heterogeneous network structure, the objective function is maximized by taking a large number of random walks along the network paths, following a pre-defined *meta-path scheme* that incorporates the node types. Dong et al. (2017) recommend that a meta-path scheme be constructed such that it begins and ends with nodes of the same type, e.g., "*User* -> *Image* -> *Annotation* -> *Image* -> *User*" in our setting. Following the meta-path scheme, random paths are generated through uniform random walks: from a user node, uniformly randomly pick an image node that is connected to the user node; then from the image node, uniformly randomly pick a connected annotation node; and so on. A neighborhood size k is also specified, such that the k nodes preceding node v and the k nodes following it are considered as the *neighborhood* of node v in Equation 5.

Evaluating the denominator in Equation 6 is computationally prohibitive for large networks. In a skip-gram model, *negative sampling* is used instead for approximation: for each v and c_t, the sigmoid function $1/(1+\exp(-X_{c_t} \cdot X_v))$ is used to calculate the probability, combined with M negative samples (i.e., nodes not in the neighborhood of v) where each negative sample u^m contributes $1/(1+\exp(X_{u^m} \cdot X_v))$ to the objective function. The model is then trained through stochastic gradient descent.

Dong et al. (2017) specify two algorithms for that share the above components but differ slightly on negative sampling. The first, called *metapath2vec*, draws negative samples randomly from all nodes, while the second algorithm, called *metapath2vec++*, draws negative samples corresponding to node c_t only from the nodes that are of the same type as c_t. Intuitively, optimizing the objective function will push the embedding vector of a negative sample node away from the embedding vector of the focal node. Thus the different negative sampling approaches will result in different spatial layouts of embedding vectors by node types. Specifically, by only considering nodes of the same type as the context node (thus not of the same type as the focal node) for the negative sample, *metapath2vec++* is likely to result in tighter clustering of the same type of nodes and more separation by type than *metapath2vec*.

Both *LINE* and *metapath2vec(++)* are effective at mapping nodes in a large-scale network into lower-dimensional vectors while preserving the structural and semantic information, and are a good fit for analyzing social curation. Compared with *LINE*, *metapath2vec(++)* has the advantage of explicitly taking into account the different node types, a crucial aspect of the semantic information of the heterogeneous network in our setting. Furthermore, *LINE* explicitly accounts for only first- and second-order proximities, while higher order proximities are preserved indirectly through them. In contrast, by using the skip-gram model with a larger neighborhood size, *metapath2vec(++)* would bring nodes that are several hops away directly into the objective function. Considering these, we use *metapath2vec(++)* as the main network embedding method for our analysis.

4.4 Deriving insights from embedding results

Embedding yields a vector representation for each node in the network. Since users, images, and annotation words are mapped into the same vector space, the embedding vectors can be used to generate important insights about consumer preferences and image characteristics. To begin, one important task when analyzing content curation is identifying *user interest groups* and *image clusters*. Well-defined distance measures among embedding vectors, e.g., the Euclidean distance, combined with clustering methods such as *K*-means, make it straightforward to separate users and images into different groups. Since the embedding vector preserves the original information of curation actions, users that are grouped into the same cluster this way would have similar preferences and interests, and images in the same cluster would have similar characteristics. Next, we can also assess the *matching between user* *interests and image characteristics*. Just like users in the same cluster would have common interests, a user group being close to an image cluster would indicate that users in that group are in general interested in the images in that cluster. The frequency of curation actions taken by users in the user cluster of images in the image cluster can also be used to corroborate such matching. Furthermore, in the embedding space, annotation words close to a user or image cluster *provide meaningful descriptions* of the common interests of these users or the common characteristics of the images. We can also account for the frequency of the association to annotation words, e.g. by counting the number of users in a cluster that are connected to the annotation word in the network. Using such frequency information, word clouds can be created to give rich descriptions of user interest groups and image clusters.

In addition to clustering by distance, images can also be grouped based on their brand association. Analyzing the embedding vectors of images of brands can reveal valuable insights on *brand perceptions and positioning*, e.g., about different user groups' interest in different brands, and about the relative positioning of different brands which can inform on the competitive market structure. Finally, platform and brand managers seek to not only understand user and image groupings, but also *predict* users' future curation actions. Embedding vectors can be used for these predictions. For example, one can predict that the closer an image is to a user in the embedding space, the more likely the user will curate the image. Alternatively, probabilistic models can be developed based on the distance between nodes, which would enable the development of recommender systems to enhance the platform's functionality.

Curation actions are taken by users to serve their personal interests or goals. These actions contain valuable information about latent consumer preferences that are important in shaping consumer demand for products. The method and tools discussed above create a holistic view that captures the global information resulting from users' curation process, and extract and present the information in a meaningful and actionable manner to researchers and managers. This is the key benefit of our proposed approach.

5. Results

We perform heterogeneous network embedding on the heterogeneous user-image-annotation network, using both *metapath2vec* and *metapath2vec++* methods. We use "*User -> Image -> Annotation -> Image -> User*" (*UIAIU*) as the main meta-path scheme, and evaluate alternative meta-path schemes. Originating from each user in the dataset, we generate 50 random paths each with a length of 100 nodes. We use a neighborhood size of 7 and 128 dimensions for the embedding space.[15] This yields a 128-dimension vector representation in the same vector space for each user, image, and annotation in our dataset. We use the first 8 weeks of data for embedding, and the remaining data for holdout evaluation.

5.1 Predictive performance

Both the heterogeneous network representation and the network embedding method are new to the marketing literature and are subject to validation. Is the heterogeneous network an effective representation that indeed captures important information about social curation? Is network embedding effective in extracting such information from the network? These are the key questions for assessing the effectiveness of our proposed approach, and an objective and convincing way to answer these questions is by evaluating the methods' predictive performance. A key task for analyzing social curation is to predict who will curate what content in the future. Since content is typically unstructured data, and little information is known about users besides their curation actions, predicting which content a user will curate in the holdout period is a challenging task, and is appropriate for evaluating our proposed approach against prevailing benchmarks.

We predict the images a user will pin in the holdout period using a k-nearest-neighbor (k-NN) approach, based directly on the embedding vectors.[16] For each user, we first identify the images she has pinned in the training period. We then identify N images that have not been pinned by the user that are closest to these previously pinned images, based on the distances between their embedding vectors. We treat those N images as the predicted images that the user would curate next. We evaluate the hit ratio

[15] A power of 2 is typically used for dimensionality in network embedding.
[16] Methods more sophisticated than k-NN may further improve predictive accuracy. However, since the goal here is to evaluate the effectiveness of our proposed approach on preserving and extracting information, rather than to optimize predictive performance per se, we simply use the k-NN method to generate prediction directly based on the embedding result.

for the first 1, 2, and 4 weeks of the holdout period, calculating the success rate based on whether the user actually pinned the images in the specified time window. We choose N=5, 10, and 20 for evaluation, and consider an instance as successful (i.e., a hit) if in the specified time window the user pinned at least one of the N recommended images.

5.1.1 Benchmark methods

Since there are thousands of images, each with limited observed characteristics, standard choice models using images as products would not work well and would not serve as a credible benchmark. Instead, we compare the predictive performance against four other popular and representative benchmark methods, ranging from social network-based techniques to various models used in the recommender systems. The first benchmark we compare to is the traditional *collaborative filtering* method, which we denote as *CF*. This is a standard and popular approach in industry, particularly in recommender systems. The second benchmark, *matrix factorization*, denoted as *MF*, is a state-of-the-art extension to the collaborative filtering method, and can learn latent low-dimensional factors underlying user-item interactions. The third benchmark method is *social network collaborative filtering*, denoted as *SNCF*, which leverages information on social connections among users and the concept of homophily. The fourth benchmark is the *community detection* method, denoted as *CD,* which partitions network nodes into groups to optimize in-group similarity. The technical details of these benchmark methods are discussed in the Online Appendix TA.1. These benchmark methods originated from different fields of research, and are commonly used in a wide range of industry settings. Comparing our method to these state-of-the-art benchmarks helps us validate the effectiveness of our proposed approach.

For our proposed approach, we evaluate all three variants as discussed in Section 4.3: the *LINE* method which performs homogeneous embedding, and the *metapath2vec* and *metapath2vec++* methods which perform heterogeneous network embeddings. Comparing these three variants with the benchmarks will help further establish the effectiveness of our proposed approach. Furthermore, comparing *metapath2vec* and *metapath2vec++* with *LINE* will also help evaluate the importance of capturing the node heterogeneity in the information network. Finally, comparing *metapath2vec* and *metapath2vec++* will help evaluate the performance implications of various embedding configurations.

5.1.2 Prediction results

[Insert Table 3 About Here]

The hit ratios, across all 108,123 users in the training dataset for all three time windows and the three levels of selectiveness (*N*=5, 10, 20) is reported in Table 3. We first note that the last benchmark, the community detection (*CD*) method, has the best performance among all the benchmarks.[17] As *CD* is the only benchmark that uses a network representation for the images, this is the first evidence that the network representation is effective in capturing information pursuant to consumers' content curation. Both *CF* and *MF* methods significantly underperform the *CD* method, showing that the collaborative filtering approach, even the state-of-the-art one that involves sophisticated matrix transformations, is not as effective as the network approach. Meanwhile, the *SNCF* method also underperforms the *CD* method (see footnote 16). This suggests that the network constructed from consumers' curation actions provides more useful information about their interests than does the following-follower social network. Interestingly, this also corroborates the finding from industry research that the conversion rate of Pinterest traffic is 22% more than that of Facebook and the consumers spend 60% more.[18]

More importantly, across all time windows and levels of selectiveness, all three network embedding methods achieve significantly better predictive performance than all the benchmark models.[19] The hit ratios of the network embedding methods are roughly twice those of *CD*, the best performing benchmark. This indicates that embedding is more effective at extracting and preserving the structural and semantic information about networks than community detection, which groups nodes

[17] The social network method *SNCF* is evaluated on a much smaller sample of 4,101 users. The social network data was downloaded through Pinterest API, the capacity constraint of which together with the large size of the social network prohibited us from obtaining the data for more users. The hit ratio for this sample is higher, likely because these are more active users. The comparable performance of the *SNCF* method is much lower than the community detection (*CD*) method. For t=1 and N=20, for example, the *CD* method's hit ratio is 0.0122 for this subset of users, six times that of *SNCF*. Detailed performance measures for this subset of users are available upon request.

[18] "Facebook vs. Pinterest: You're Investing, But What Are Your Goals?" URL: https://www.bloomreach.com/en/blog/2013/04/facebook-vs-pinterest-youre-investing-but-what-are-your-goals.html, accessed in December 2018.

[19] The hit ratios are low across all methods. This simply reflects the difficulty of the prediction task: we are selecting a few images out of a total of more than one-thousand, and each user, on average, pins fewer than one image in the holdout period. For comparison, the hit ratio of a naïve recommender, which randomly recommends *N* images, is about *one order of magnitude lower* than those achieved by the network embedding methods.

into subsets based directly on the connections. These performance comparisons, between the proposed and benchmark methods and between the different benchmarks, together confirm the importance of both the heterogeneous network representation and the network embedding method. By effectively incorporating both components, our proposed approach delivers significantly better performances than other state-of-the-art benchmarks.

Among the network embedding approaches, *metapath2vec++* has similar predictive performance to *LINE*, while *metapath2vec* has slightly worse performance. The importance of recognizing node heterogeneity thus is not obvious in this comparison. However, an important advantage of the heterogeneous embedding method is that researchers can specify different meta-path schemes to emphasize different connections among the entities. For predicting users' pinning of images, emphasizing the user-image connections in the meta-path scheme (e.g., using *UIULAIU* or *UIUIULAIU* instead of the original *ULAIU* scheme) can be expected to improve the performance. In contrast, emphasizing other connections, e.g., *ULALAIU*, may deliver worse performance. To validate this, we also perform predictions using alternative meta-path schemes for *metapath2vec* and *metapath2vec++*. The result is reported in Table 4.

[Insert Table 4 About Here]

As clearly shown in the table, varying meta-path schemes has a significant impact on the predictive performance. As conjectured, emphasizing user-image connections in the meta-path scheme leads to better predictive performance, whereas emphasizing other connections, thus comparatively de-emphasizing user-image connections, leads to worse predictive performance. More importantly, by emphasizing the user-image connections, using both *UIULAIU* and *UIUIULAIU* schemes for *metapath2vec* and *metapath2vec++* generates significantly better prediction accuracy than using *LINE*. This shows that recognizing node heterogeneity is indeed important, and heterogeneous embedding extracts more information than homogeneous embedding.

Table 5 reports a more detailed comparison between *LINE* and *metapath2vec* as well as *metapath2vec++* (using the original *ULAIU* meta-path scheme), by evaluating the performance based on the activeness of users. The table shows that *LINE* performs better for users who pinned two or more images in the training period. For the great majority of users who pinned only one image (94% of all users), however, both *metapath2vec* and *metapath2vec++* outperform *LINE*. Thus when only limited information is available, the heterogeneous embedding method outperforms homogeneous embedding, possibly because explicitly accounting for the semantic information of node types makes information extraction more efficient. Data sparsity is a key challenge confronting business managers. This comparison suggests that heterogeneous embedding is more effective than homogeneous embedding for addressing this challenge.

[Insert Table 5 About Here]

Two important implications can be drawn from our analysis of the methods' predictive performance. First, the comparison clearly shows that heterogeneous network representation effectively captures key information about social curation. Second, the network embedding method is effective in extracting this information, and compared with homogeneous embedding, heterogeneous embedding can achieve significantly better predictive performance, and is more effective when only a little information is available about users. Combining both network representation and network embedding delivers significantly better performance than other extant methods, which confirms the effectiveness of our proposed approach.

5.2 Interpreting the embedding results

The information captured through the network representation and extracted through embedding not only can be used for prediction, but also can reveal insights for marketing managers. We now discuss the various ways to use the embedding results to shed light on consumers, images, and brands. A discussion on visualizing the raw embedding vectors is available in the Online Appendix (TA.2). In the discussion below, we analyze the embeddings generated by *metapah2vec* using the *ULAIU* meta-path scheme. Embeddings generated by *metapath2vec++* and using other meta-path schemes can be interpreted in a similar fashion.

5.2.1 Consumer interest groups and image clusters

An important goal of analyzing social curation is to identify consumers who share similar interests and images with similar characteristics, and to match consumers with images they are interested in. These user interest groups and image clusters can be discerned from the embedding vectors through clustering analysis.

We perform constrained K-means clustering of the embedding vectors (Bradley, et al. 2000).[20] Using the silhouette value (Peter, 1987) as the criteria of selecting the optimal K, the analysis yields 15 user clusters and 8 image clusters. The number of users in each cluster and their average characteristics are reported in Table 6.[21] The clusters differ in several aspects. First, the size of the clusters varies significantly, with the number of users in a cluster ranging from 295 for cluster 15 to 16,314 for cluster 10. Secondly, their level of activity also differs. Judging by the number of pins of the images in our dataset, cluster 15 has 35% more activity than cluster 9, while judging by the total number of pins, cluster 12 is 3.4 times more active than cluster 1.[22] They also differ in the extent of their social connections with cluster 14 having more than eight times the number of followers as cluster 1. The activity levels, social connections, and tenure have only a moderate correlation, suggesting that the information reflected from these cluster-level characteristics goes beyond a general level of activeness (i.e., more active users would pin more images and have denser social networks). Taken together, clusters 11 and 14 demonstrate the properties of thoughtful curators, with a high number of followers and moderately high activity; clusters 4 and 12 can be described as active users, with a high number of pins but not highly followed; clusters 1 and 2 are comparatively less active, with fewer pins and followers.

[Insert Table 6 About Here]

Similarly, the number of images and the average characteristics of the image clusters are reported in Table 7 and Table 8. Two of those clusters are large, with clusters 8 and 4 having 542 and 184 images, respectively. They also contain the most popular images (i.e., those pinned frequently by users). Clusters 2 and 3, in contrast, have much lower pin activity. The number of pin actions is not determined by brands. Cluster 7 has the second highest percentage of branded images but few pin

[20] Each image cluster is constrained to have at least 50 images, to avoid focusing on idiosyncrasies.
[21] Note that the clustering is performed solely based on the embedding vectors, not the observed consumer characteristics, which are summarized on a post hoc basis.
[22] We note that users in our dataset all have pinned at least one image in the dataset. The difference in pin activity is thus higher if this condition is taken into account.

actions, while cluster 8 has high average pin actions even though the percentage of branded images is low.

Images in different clusters also vary significantly in their pixel characteristics, as shown in Table 8. While these image characteristics extracted using OpenCV and Google Vision API are not used directly in network embedding and clustering, we can use them on a post hoc basis to gain additional understanding of the image clusters. Relating these characteristics to pin activity, we can see that images of portrait format, having lighter colors, and having a lower intensity are more popular. Images with higher hue and saturation also garner more pins. For example, images in cluster 4 are pinned most frequently, and they have the standard height-to-width ratio for portrait orientation, and have the lightest color and lowest average intensity. In contrast, images in clusters 2 and 7, the least popular by pin activity, are wider and have the lowest hue and saturation.

[Insert Tables 7-9 About Here]

5.2.2 Matching consumers with images

It is important to not just identify user interest groups and image clusters, but to also discern the interrelationship between them. Table 10 reports the average number of pins between each user and image cluster for the training period. The number of pins differs considerably across user-image cluster pairs, clearly indicating that different user interest groups prefer different groups of images. While some of the difference can be attributed to popularity (e.g., image clusters 4 and 8 in general are pinned more than the other clusters), most cannot. Thus, insights can be drawn from different angles. First, image cluster 8 has broad appeal, as many user groups have moderately high pin counts for this image cluster. In contrast, image cluster 4 has a much narrower targeted appeal, being pinned mainly by four groups of users (user clusters 1, 2, 6, and 7). Second, some user groups, such as clusters 1 and 2, have focused interest, pinning mostly images from only image cluster 4, whereas other user groups, such as clusters 13 and 14, actively pin images from multiple image clusters. Finally, different user groups clearly differ in their preferred images. User clusters 1 and 2, for example, actively pinned images in image cluster 4 but hardly any in image cluster 8, while user clusters 3, 4, and 5 did the opposite, pinning mostly image cluster 8 but not 4. User cluster 13 had a significant number of pins in image cluster 5, while other user clusters had few pins. Thus, meaningful matches can be made between user groups and image clusters, indicating horizontal differentiation by taste rather than vertical differentiation by quality. More importantly, this match is highly stable for the holdout period: the correlation between the in-sample and out-of-sample average pins is a high 0.66. (The average pin numbers for the holdout period are reported in Table TA.3.2 in the Online Appendix.) Thus, this match likely reflects the underlying interests and characteristics at a fundamental level.

5.2.3 Understanding consumers and images through curation words

The clusters help identify user interest groups, image clusters, and the matching between them. Curation words can further shed lights on the nature of these interests and characteristics. Based on the distances between the vectors in the embedding space, the user and image clusters differ significantly in the set of words with which they are associated. For example, image cluster 4 is close to words such as *garden*, *backyard*, *patio*, and *outdoor*, suggesting it is likely related to outdoor furniture. Image cluster 3 is close to words such as *logo*, *lamin*, *sketch*, and *craft*, suggesting these images are more design and idea oriented. Similarly, user cluster 2 is interested in outdoor furniture, being associated with *backyard*, while user clusters 3, 4, and 5 are interested in indoor furniture, being close to *interior*. The 10 curation words closest to each image and user cluster are available in the Online Appendix (Tables TA.3.3 and TA.3.4).

[Insert Figures 5 and 6 About Here]

Accounting for the frequency of curation words can provide additional interpretation. For each user and image cluster, we count the number of users and images in the cluster that are connected to each curation word, and create the word cloud[23]. Figures 5 and 6 show these word clouds for the user and image clusters, respectively. As shown in the word clouds, the brand name Ashley is salient for image cluster 2; image cluster 4 has a balanced representation of outdoor related curation words; image cluster 7 emphasizes Scandinavian and modularity, suggesting the images are likely related to IKEA even though the brand name itself is not salient. Similarly, user cluster 2 reveals a clear interest in outdoor furniture; user cluster 4 pays attention to coloring, with words like *blue*, *white*, and *paint* being

[23] We use the tool *wordle* to create the word cloud. http://www.wordle.net/

salient; user cluster 9 seems to focus on creative design, with representative words such as *dream*, *craft*, and *idea*.

Taken together, curation words provide rich descriptions to extract insights about users and images. While the clusters show the existence of user interest groups and image clusters and their alignment, curation words reveal the nature and meaning of user interests and image characteristics. Generating these rich descriptions by analyzing curation words is what makes studying social curation unique, as it sheds light on the underlying reason for collecting the images.

5.2.4 Analyzing brands

Brand managers are keenly interested in reaching consumers through social curation platforms. Understanding brand positioning and finding interested consumers are the keys to success. As discussed in Section 3, using Amazon Mechanical Turk, a significant portion (66%) of the images in our dataset are identified as belonging to ten popular furniture brands. Using the distances between embedding vectors, we can draw insights related to brands, similar to assessing the match between user and image clusters as discussed in section 5.2.2. The average distance between each user cluster and each brand is available in the Online Appendix (Table TA.3.5).

Across the brands, Crate & Barrel and Restoration Hardware are the closest to users, while IKEA and La-Z-Boy are the farthest. Across the users, cluster 2 is the closest to the brands in general, while cluster 9 is the farthest. This distance is indicative of users' interest in each brand: the correlation between the distance of a user cluster to a brand, and the average number of images in the corresponding brand that was pinned by users in that user cluster is -0.11, (i.e., the closer a user cluster is to a brand, the more likely those users will pin images of that brand). This relationship is also stable for the holdout period, where the corresponding correlation is -0.22. More interestingly, this relationship extends to branded images outside our embedding dataset: for the images in the holdout sample (i.e., the images for which the first pin occurred after the first 8 weeks) that are also associated with one of the brands, the average pins of them by each user cluster is also negatively correlated with the user-brand distance, and the correlation is -0.32. This test of using holdout images in the holdout period shows that the distance between user clusters and brands reflects the underlying match between user interests and brand characteristics, which transcends individual images. This is an important finding for brand managers who can target the interested user groups when new images of their brands become available.

The brands' relative positioning can also be assessed through embedding results. Figure 7 presents a 2-D plot of the relative positioning among brands using multi-dimensional scaling (MDS), based on the average distance between the embedding vectors of branded images. The figure shows that Ethan Allen and Crate & Barrel occupy a central position in the map, while La-Z-Boy and Ashley are at peripheral positions. Such positioning maps are of interest to brand managers who seek to identify similar and competing brands.

[Insert Table 10 and Figure 7 About Here]

The embedding results can also be used to assess the level of "cohesion" of each brand. Intuitively, if images belonging to a brand tend to have a consistent style, then their embedding vectors would be similar, whereas if the brand is less cohesive, then the embedding vectors would be more dispersed. We report a measure of cohesion using the average distance between the embedding vector of each branded image to the corresponding brand centroid. As Table 10 shows, Crate & Barrel and Restoration Hardware have a relatively high level of cohesion with the lowest average distance to the brand centers. In contrast, IKEA has the most dispersed images and thus the lowest cohesion. This is understandable since the former brands have distinct styles, while IKEA is a more general-purpose brand that crosses many furniture categories. The ability to reveal this information from embedding results further validates the effectiveness of the proposed approach.

5.3 Managerial discussions

On social curation platforms such as Pinterest, the curation network is formed organically, and contains valuable information on consumer preferences and image characteristics, and by extension on brands. Our proposed approach can be used by brand managers in several ways. First, the alignment between users, images, and annotation words provides insights on what consumers are interested in which products and for what purposes. Compared with traditional demographics-based consumer segments, the user groups inferred from our analysis would share interests at a more fundamental level. Compared with latent classes of consumers identified using a limited number of coefficients in choice models, these user groups would share broader and more versatile interests. By identifying and describing interest groups, our approach can help brand managers better design their content marketing, align products with consumers' interests, and improve the effectiveness of contextual targeting. Furthermore, the heterogeneous network provides brands a global and holistic view on brand positioning in the competitive landscape, and managers can study the network structure of images and annotations to analyze new consumer-defined trends, discover potential demands, and identify new consumer segments. Finally, our study not only presents an analytical approach, but also provides a rich set of ready-to-use tools for implementation. The methods used in our study for prediction, clustering analysis, text analysis, brand positioning, and brand cohesion measurement, with their straightforward implementations, can be directly applied by managers to real-world settings to help improve business performance.

6 Conclusion

Social curation, the discovery, collection, organization, and annotation of the vast content on the Internet by millions of consumers in a collaborative manner, is a rapidly growing industry trend. Besides Pinterest, many other social curation tools and sites exist (e.g., Storify and Pearltrees). Established social media channels such as YouTube and Instagram have also added social curation functions to their sites. E-commerce firms have also participated in this trend. For example, Amazon has a Pinterest-like feature called "Interesting Finds," which offers shoppers a curated feed of products. This integration of social curation and e-commerce calls for research to understand the social curation process, and harness the information on social curation platforms for marketing researchers and practitioners. Such information is especially valuable, since curation actions are taken by consumers and reflect their preferences and interests. Brands that can better leverage social curation to quickly identify interested consumers and bring matching products to them will have more success in this increasingly competitive landscape.

Despite the exciting new possibilities, harnessing the information from social curation is challenging. Due to the large-scale and unstructured data, standard choice models are not effective. The novel approach proposed in this study, to first represent social curation using a heterogeneous network, and then use the heterogeneous network embedding method to extract the information into a lower-dimensional vector space for further analysis, is an effective way to address this challenge. This proposed approach enables a holistic view of users' collective curation actions, while accounting for the interconnections among users, images, and curation words in an elegant manner. Compared with extant non-network and network-based state-of-the-art benchmarks, our proposed approach achieves significantly better predictive performance. Our proposed approach also yields rich insights. We can identify a multitude of consumer interest groups and image clusters each with distinct characteristics, and evaluate the match between consumer and image groups. Consumers' interests and image characteristics can also be interpreted using annotation words. Leveraging brand information, we further assess the brand positioning both visually and analytically. All these insights contribute to our understanding of the social curation phenomenon, and our proposed approach provides tools and actionable intelligence to marketing managers.

While this study focuses on crowd-based social curation platforms, our approach of using a heterogeneous network to represent a situation and using network embedding to extract and analyze the information has good potentials of generalization. On e-commerce websites, a purchase network can be constructed, where consumers and products are represented as nodes, and purchases form the edges. On a product review platform, the ratings and reviews given by consumers can be treated as edges in a network of products and consumers. In a paid search setting, a heterogeneous network can be constructed with users, keywords, and firm websites according to the search and visit activity. On image and video sites such as Instagram and YouTube, the viewing, tagging, and commenting behaviors can also be modeled with networks. A network can be used to represent consumers, items, and purchases even in the traditional retail store context, where consumers purchase grocery items, which have usually been analyzed using choice models. Once a network representation is constructed, the network embedding method can be used to extract and analyze the information. Thus, our proposed approach can be generalized to many marketing situations beyond social curation, creating new research opportunities on important substantive issues. This proposed approach is the key methodological contribution of our study.

As the first step towards exploring the value of social curation, network representation, and network embedding, our study opens up avenues for future research. First, due to methodological limitation, we focus on the static analysis of curation. A natural extension would investigate the dynamic curation process and study the trajectory of content propagation. However, a dynamic embedding method is needed for such analysis, which we leave for future study. Second, the heterogeneous network in our study is constructed solely based on curation actions, while various user and image characteristics are included for descriptive analysis on a post hoc basis. More advanced network embedding approaches may explicitly take into account node attributes when generating the embedding vectors, a topic we also leave for future study. Third, while our study provides a descriptive view of social curation and insights are derived based on correlations, future study can focus on specific aspects of the curation process to reveal causal findings through field experiments. Finally, our study focuses on understanding the early stage of consumer needs and wants, while future research can associate consumer-expressed interests in social curation with actual purchases.

References

Anderson, E.W. (1998), "Customer satisfaction and word of mouth," *Journal of Service Research*, 1(1), 5–17.

Berger, J. and K.L. Milkman, (2012), "What makes online content viral?" *Journal of Marketing Research*, 49(2), 192–205.

Berger, J. and E. Schwartz, (2011), "What drives immediate and ongoing word of mouth?" *Journal of Marketing Research*, 48(5), 869–880.

Bradley, P. S., Bennett, K. Pl, and Demiriz, A. (2000). Constrained K-Means Clustering. MSR-TR-2000-65, Microsoft Research.

Cao, Shaosheng, Wei Lu, and Qiongkai Xu. 2015. Grarep: Learning graph representations with global structural information. In *CIKM*. ACM, 891–900.

Chang, S. W. Han, J. Tang, G.-J. Qi, C. C. Aggarwal, and T. S. Huang. Heterogeneous network embedding via deep architectures. In *KDD*. ACM, 2015.

Chen, Mo, Qiong Yang, and Xiaoou Tang. 2007. Directed Graph Embedding. In *IJCAI*. 2707–2712.

Chevalier, Judith and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (August), 345-54.

Chintagunta, P.K., S. Gopinath, S. Venkataraman, (2010), "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets," *Marketing Science*, 29(5), 944–957.

Chu, Junhong and Puneet Manchanda (2016),"Quantifying Cross and Direct Network Effects in Online Consumer-to-Consumer Platforms." *Marketing Science*, 35:6 , 870-893

Daixin Wang, Peng Cui, and Wenwu Zhu. 2016. Structural deep network embedding. In *KDD*. ACM, 1225–1234.

Defferrard, Michael, Bresson, Xavier, and Vandergheynst, Pierre (2016). "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," In *NIPS*.

Dong, Y., N. V. Chawla, and A. Swami (2017), "metapath2vec: Scalable Representation Learning for Heterogeneous Networks."

Dong, Yuxiao, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In *KDD*.

Gao, Hongchang and Heng Huang (2018), "Deep Attributed Network Embedding," In *IJCAI*, pp. 3364 – 3370.

Godes, D. and D. Mayzlin, (2004), "Using online conversations to study word-of-mouth communication," *Marketing Science*, 23(4), 545–560.

Godes, D. and D. Mayzlin, (2009), "Firm-created word-of-mouth communication: Evidence from a field test," *Marketing Science*, 28(4), 721–739.

Godes D. and J. Silva, (2012), "Sequential and temporal dynamics of online opinion," *Marketing Science*, 31(3), 448–473.

Grover, Aditya and Jure Leskovec, (2016), "node2vec: Scalable feature learning for networks," In *KDD*. ACM, 855–864.

Hall, C., and M. Zarro, (2012), "Social curation on the website pinterest.com," *Proceedings of the American Society for Information Science and Technology*, 49, 1–9.

Hennig-Thurau, T., K.P. Gwinner, G. Walsh, D.D. Gremler, (2004), "Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?" *Journal of Interactive Marketing*, 18(1), 38–52.

Hill S., F. Provost, and C. Volinsky (2006), "Network-based marketing: Identifying likely adopters via consumer networks," *Statist. Sci.* 21(2):256–276.

Katona, Z. and M. Sarvary (2008). "Network Formation and the Structure of the Commercial World Wide Web." *Marketing Science*, 27:5 , 764-778.

Kipf, T.N. and M Welling (2017), "Semi-supervised classification with graph convolutional networks," *5th International Conference on Learning Representations* (ICLR-17).

Kumar, V. and K. Sudhir, (2019), "Can Friends Seed More Buzz and Adoption?" working paper.

Li, Jundong, Harsh Dani, Xia Hu, Jiliang Tang, Yi Chang, and Huan Liu, (2017), "Attributed network embedding for learning in a dynamic environment," In *2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*. ACM, New York, NY, USA, 387-396, 2017.

Li, X., M. Shi, and X. Wang, (2019), "Video mining: Measuring visual information using automatic methods," *International Journal of Research in Marketing*, https://doi.org/10.1016/j.ijresmar.2019.02.004

Li, X. and L.M. Hitt, (2008), "Self-selection and information role of online product reviews," *Information Systems Research*, 19(4), 456–474.

Liu, L., D. Dzyabura and N. Mizik (2017), Visual Listening In: Extracting Brand Image Portrayed on Social Media. working paper.

Liu, Ninghao, Xiao Huang, and Xia Hu, (2017), "Accelerated local anomaly detection via resolving attributed networks," *26th International Joint Conference on Artificial Intelligence (IJCAI'17)*, Carles Sierra (Ed.). AAAI Press 2337-1343.

Lu, Y., P.V. Singh, and B. Sun (2017), "Is Core-Periphery Network Good for Knowledge Sharing? A Structural Model of Endogenous Network Formation on a Crowdsourced Customer Support Forum." *Management Information Systems Quarterly*, 41(2), 607-628.

Ma, L., B. Sun and S. Kekre, (2015), "The Squeaky Wheel Gets the Grease - An Empirical Analysis of Customer Voice and Firm Intervention on Twitter," *Marketing Science*, Vol. 34, No. 5, pp. 627-645.

Miha Grčar, Dunja Mladenič, Blaž Fortuna, and Marko Grobelnik. 2005. "Data sparsity issues in the collaborative filtering framework," In *WebKDD*, Berlin, Heidelberg, 58-76.

Miller, H. J., Chang, S., & Terveen L. G. (2015). I love this site! vs. it's a little girly: Perceptions of and initial user experience with Pinterest. In *CSCW*. pp. 1728–1740.

Moe W.W. and D.A. Schweidel, (2012), "Online product opinion: Incidence, evaluation and evolution," *Marketing Science*, 31(3), 372–386.

Moe W.W. and M. Trusov, (2011), "The value of social dynamics in online product ratings forums," *Journal of Marketing Research*, 48(3), 444–456.

Perozzi, B., A. Rami and S. Skiena. (2014), "Deepwalk: Online learning of social representations," In *KDD*. 701–710.

Rousseeuw, P.J., (1987) "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Computational and Applied Mathematics*. 20: 53-65.

Phan, T.Q. and D. Godes, (2018), "The Evolution of Influence through Endogenous Link Formation," *Marketing Science,* Forthcoming.

Ribeiro, Leonardo F.R., Pedro H.P. Saverese, and Daniel R. Figueiredo (2017), "struc2vec: Learning Node Representations from Structural Identity," *23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*. ACM, New York, NY, USA, 385-394.

Roweis, S.T. and L.K. Saul, (2000), "Nonlinear dimensionality reduction by locally linear embedding," *Science*, Vol. 290, Issue 5500, pp. 2323-2326.

Schlosser, A.E., (2005), "Posting versus lurking: Communicating in a multiple audience context," *Journal of Consumer Research*, 32(2), 260–265.

Schweidel D. and W.W. Moe, (2014), "Listening in on social media: A joint model of sentiment and venue format choice," *Journal of Marketing Research*, 51(4), 387–402.

Stephen, A.T. and J. Galak, (2012), "The effects of traditional and social earned media on sales: A study of a microlending marketplace," *Journal of Marketing Research*, 49(5), 624–639.

Stephen, A.T. and O. Toubia (2010), "Deriving Value from Social Commerce Networks," *Journal of Marketing Research* 47, no. 2: 215-228.

Tang, J., M. Qu, M. Wang, M. Zhang, J. Yan and Q. Mei, (2015), "LINE: Large-scale information network embedding." In *WWW*, pp. 1067–1077, 2015.

Tirunillai, S. and G.J. Tellis, (2012), "Does chatter really matter? Dynamics of user-generated content and stock performance," *Marketing Science*, 31(2), 198–215.

Tirunillai, S. and G.J. Tellis, (2014), "Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation," *Journal of Marketing Research*, 51(4), 463–479.

Tong, H., S. Papadimitriou, J. Sun, P.S. Yu and C. Faloutsos, (2008), "Colibri: fast mining of large static and dynamic graphs," In *KDD*. 686–694.

Toubia O. and A.T. Stephen, (2013), "Intrinsic vs. image-related utility in social media: Why do people contribute content to Twitter?" *Marketing Science*, 32(3), 368–392.

Trusov, M., R.E. Bucklin, K. Pauwels, (2009), “Effects of word-of-mouth versus traditional marketing: Findings from an Internet social networking site,” *Journal of Marketing*, 73(5), 90–102.

Wang, Chi, Rajat Raina, David Fong, Ding Zhou, Jiawei Han and Greg Badros (2011). “Learning Relevance from Heterogeneous Social Network and Its Application on Online Targeting.” *34th International ACM SIGIR conference on Research and development in Information Retrieval (SIGIR’11).* ACM, New York, NY, USA, 655-664.

Wang, Suhang, Charu Aggarwal, Jiliang Tang, and Huan Liu. 2017. Attributed Signed Network Embedding. In *CIKM*. ACM, New York, NY, USA, 137-146.

Xiao Huang, Jundong Li, and Xia Hu. Label informed attributed network embedding. In *WSDM*, 2017.

Yang, Zhilin, William Cohen, and Ruslan Salakhudinov. 2016. Revisiting Semi-Supervised Learning with Graph Embeddings. In *ICML*. 40–48.

Zhang, Shunyuan, Dokyun Lee, Param Vir Singh, Kannan Srinivasan, (2017), “How Much Is An Image Worth? An Empirical Analysis of Property’s Image Aesthetic Quality on Demand at AirBNB.” Working paper.

Zhang, Yuchi and David Godes, "Learning from Online Social Ties," *Marketing Science,* Forthcoming.

Zhong, Changtao, Dmytro Karamshuk and Nishanth Sastry (2015). “Predicting Pinterest: Automating a distributed human computation.” In *WWW*. 1417-1426.

Zhu, Shenghuo, Kai Yu, Yun Chi, and Yihong Gong. 2007. Combining content and link for classification using matrix factorization. In *SIGIR*. 487–494.

Tables and Figures

Table 1: User Characteristics

	Mean	SD	Min	Max
Num of Pins of Furniture Images	1.12	0.55	1	20
Num of Total Pins	9748.29	19586.90	1	308751
Num of Followings	365.88	2254.23	0	348408
Num of Followers	1163.46	28103.09	0	4588252
Num of Pin Boards	52.73	61.07	1	3038
Tenure (years)	3.52	1.75	0	7

Table 2: Image Brands

Brand	# Images In-Sample	# Images Holdout
Williams Sonoma	39	27
American Signature	45	46
Ashley	50	63
Berkshire Hathaway	21	15
Crate & Barrel	88	46
Ethan Allen	48	32
IKEA	161	141
La-Z-Boy	19	13
Raymour Flanigan	21	22
Restoration Hardware	223	122

Table 3: Prediction Performance Comparison

Method	Week 9 (t=1)			Week 9-10 (t=2)			Week 9-12 (t=4)		
Top N =	5	10	20	5	10	20	5	10	20
Item-based CF (CF)	1.8E-05	1.8E-05	7.4E-05	3.7E-05	5.5E-05	0.00014	4.6E-05	8.3E-05	0.00023
Matrix Factorization (MF)	0.00011	0.00015	0.00018	0.00014	0.00023	0.00032	0.00027	0.00039	0.00063
Social network-based (SNCF) (*)	*0.00098*	*0.00122*	*0.00122*	*0.00146*	*0.00195*	*0.00195*	*0.00146*	*0.00195*	*0.00195*
Community detection (CD)	0.00051	0.00086	0.00179	0.00074	0.00135	0.0027	0.00107	0.00207	0.00412
Homogeneous embedding (LINE)	0.00128	0.00237	0.00316	0.00203	0.00351	0.00486	0.00308	0.00511	0.00696
Heterogeneous embedding (metapath2vec)	0.00107	0.00174	0.00304	0.00166	0.00281	0.00495	0.00232	0.00399	0.00685
Heterogeneous embedding (metapath2vec++)	0.00116	0.00218	0.00344	0.00179	0.00343	0.00542	0.00253	0.00479	0.00752

* Social-network based prediction (SNCF) is made for a much smaller subset of users due to data availability, and the hit ratios are <u>not directly comparable</u> to the other methods. Please see footnote 16 for discussion.

Table 4: Prediction Performance – Alternative Meta-path Schemes

Method	Week 9 (t=1)			Week 9-10 (t=2)			Week 9-12 (t=4)		
Top N =	5	10	20	5	10	20	5	10	20
LINE	0.00128	0.00237	0.00316	0.00203	0.00351	0.00486	0.00308	0.00511	0.00696
Metapath2vec									
UIAIU	0.00107	0.00174	0.00304	0.00166	0.00281	0.00495	0.00232	0.00399	0.00685
UIUIAIU	**0.00145**	**0.0024**	**0.00349**	**0.00224**	**0.00368**	**0.00548**	**0.003**	**0.00504**	**0.0076**
UIUIUIAIU	**0.00217**	**0.00308**	**0.00424**	**0.00331**	**0.00468**	**0.00657**	**0.00433**	**0.00659**	**0.00928**
UIAIAIAIU	0.00094	0.00182	0.00309	0.00151	0.00283	0.00487	0.0021	0.00398	0.00678
UAUAUAIAU	0.00098	0.00149	0.0025	0.0015	0.00234	0.00387	0.00215	0.00341	0.00557
Metapath2vec++									
UIAIU	0.00116	0.00218	0.00344	0.00179	0.00343	0.00542	0.00253	0.00479	0.00752
UIUIAIU	**0.00163**	**0.00289**	**0.00405**	**0.00247**	**0.00447**	**0.00622**	**0.00337**	**0.00626**	**0.00879**
UIUIUIAIU	**0.00179**	**0.0029**	**0.00414**	**0.00276**	**0.00436**	**0.0063**	**0.00378**	**0.00616**	**0.00897**
UIAIAIAIU	0.00117	0.00195	0.0033	0.0018	0.00302	0.00529	0.00252	0.00437	0.00739
UAUAUAIAU	0.00069	0.00123	0.00233	0.00117	0.00204	0.00352	0.00179	0.003	0.00523

Table 5: Prediction Performance – by User Activeness

User Pin Count	Num Users	**Method**	Week 9 (t=1)			Week 9-10 (t=2)			Week 9-12 (t=4)		
		Top N =	5	10	20	5	10	20	5	10	20
1	101721	*LINE*	0.00086	0.00158	0.00215	0.00138	0.00239	0.00334	0.00205	0.00349	0.0048
		metapath2vec	0.00087	0.00132	0.00224	0.00119	0.00202	0.0036	0.00167	0.00286	0.00499
		metapath2vec++	0.00093	0.00161	0.00254	0.00132	0.00242	0.00385	0.00184	0.00337	0.00537
2+	6402	*LINE*	0.00797	0.01484	0.01921	0.0125	0.02124	0.0289	0.01937	0.03093	0.04139
		metapath2vec	0.00437	0.00843	0.01578	0.00906	0.01546	0.0264	0.01265	0.02187	0.03639
		metapath2vec++	0.00469	0.01125	0.01781	0.00922	0.01953	0.0303	0.01359	0.02734	0.04171

Table 6: User Clusters

Cluster	#Users	#Pins	#Total Pins	#Followings	#Followers	#Boards	Tenure
1	1286	1.10	4775.07	197.34	221.26	42.07	2.95
2	13906	1.13	6873.04	248.49	563.70	48.07	3.56
3	10732	1.12	7525.96	287.42	408.48	49.12	3.73
4	2422	1.19	14807.40	496.61	941.08	66.86	3.51
5	12209	1.16	11233.22	357.84	1365.81	55.70	3.76
6	7374	1.31	8824.81	412.02	550.21	50.79	2.91
7	3936	1.11	7739.68	321.53	473.72	52.44	3.63
8	5778	1.21	14396.54	490.98	1030.92	60.73	3.91
9	1281	1.07	12361.29	438.41	714.24	48.77	3.25
10	16314	1.32	11095.67	435.10	1512.62	66.96	3.92
11	12245	1.10	10426.92	308.43	1851.56	40.94	3.58
12	1907	1.21	16305.40	550.44	955.37	66.04	3.46
13	12433	1.22	7846.57	327.87	1686.46	41.49	2.79
14	5795	1.13	10513.49	568.19	1860.52	59.86	3.35
15	295	1.44	12364.38	412.84	1277.88	75.03	4.38

Table 7: Image Clusters

Cluster	Num Images	Num Pins	% Having Brand
1	50	21.16	60.00%
2	50	2.32	60.00%
3	50	5.36	66.00%
4	184	311.09	66.85%
5	101	92.70	65.35%
6	60	58.72	75.00%
7	50	2.44	74.00%
8	542	247.94	64.76%

Table 8: Image Characteristics by Cluster

Cluster	height	width	Red	Green	blue	lightness	hue	saturation	intensity
1	407.8	369.3	152.5	140.9	132.2	150.0	54.2	68.7	161.3
2	361.1	399.5	160.5	151.5	142.2	157.3	25.9	48.8	162.1
3	565.0	438.7	165.4	156.1	144.6	163.6	43.0	62.7	173.5
4	440.7	318.3	145.1	139.2	127.5	146.2	45.0	66.8	153.7
5	424.3	381.2	158.9	151.0	143.9	157.4	44.4	52.0	163.4
6	435.0	329.8	158.6	152.1	141.2	159.4	46.0	66.7	168.9
7	385.5	461.3	187.4	180.0	175.2	185.8	32.0	41.3	192.0
8	481.7	309.4	155.7	148.4	139.3	155.4	44.2	56.8	162.2

Table 9: Average Pins by User-Image Clusters (Training)

		Image Cluster							
		1	2	3	4	5	6	7	8
User Cluster	1	0.00	0.00	0.00	55.07	0.00	0.00	0.00	0.51
	2	0.07	0.07	0.00	57.04	0.23	0.01	0.01	0.31
	3	0.00	0.00	0.00	2.57	0.19	0.02	0.00	18.42
	4	1.57	0.00	0.00	0.45	0.00	0.07	0.00	19.53
	5	0.00	0.21	0.02	3.01	0.28	0.01	0.11	18.18
	6	0.03	0.00	0.00	59.79	0.05	0.00	0.00	1.29
	7	0.00	0.00	0.00	49.94	0.00	0.00	0.00	2.41
	8	0.00	0.00	0.07	0.33	0.05	0.00	0.07	19.78
	9	0.00	0.00	0.00	0.30	0.00	0.13	0.00	13.42
	10	0.34	0.00	0.10	2.22	0.00	0.08	0.02	20.12
	11	0.00	0.02	0.00	0.17	0.09	0.00	0.02	18.81
	12	0.10	0.00	0.00	1.28	0.41	0.00	0.00	19.94
	13	0.00	0.05	0.58	3.77	31.65	0.13	0.77	12.63
	14	10.36	1.76	1.00	8.29	1.09	4.56	0.03	13.65
	15	0.00	0.00	0.00	0.55	0.00	2.82	0.00	21.86

All numbers scaled up by 10^4.

Table 10: Brand Cohesion

Brand	#Images	Avg. Distance to Center
Williams Sonoma	39	2.858
American Signature	45	2.858
Ashley	50	2.878
Berkshire Hathaway	21	2.702
Crate & Barrel	88	2.599
Ethan Allen	48	2.694
IKEA	161	3.072
La-Z-Boy	19	2.751
Raymour Flanigan	21	2.668
Restoration Hardware	223	2.589

Figure 1: Example of a user page and a pin board on Pinterest

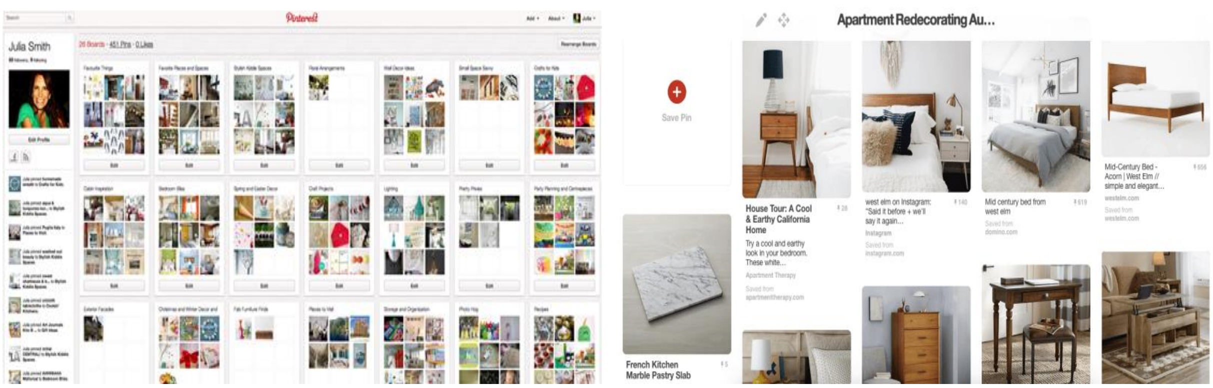


Figure 2: Histogram of the Number of Pin Actions by Image

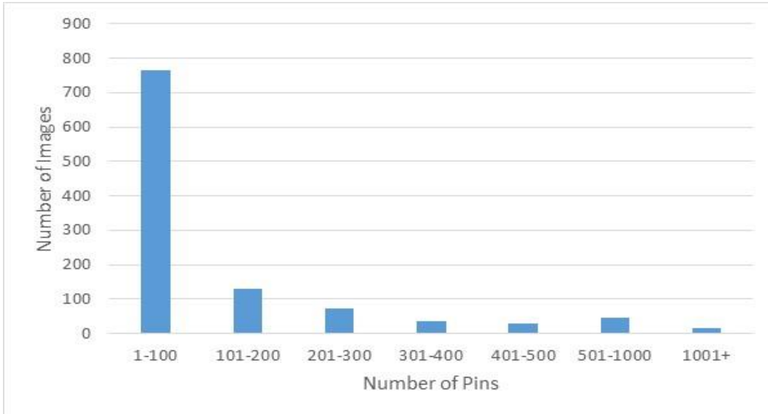


Figure 3: Sample Images in the Dataset

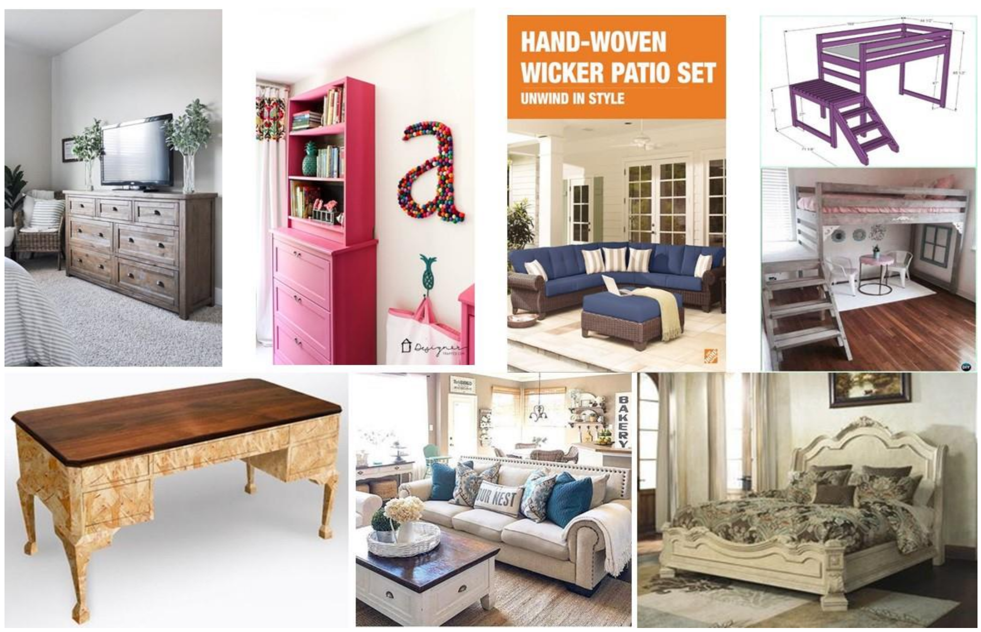


Figure 4: Heterogeneous Network Representation and Embedding

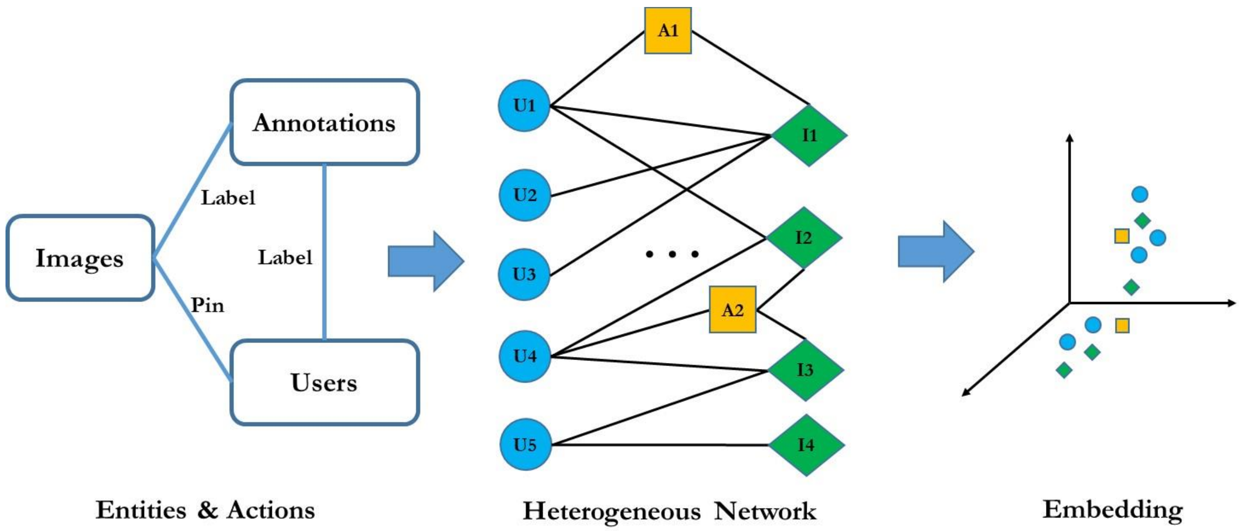
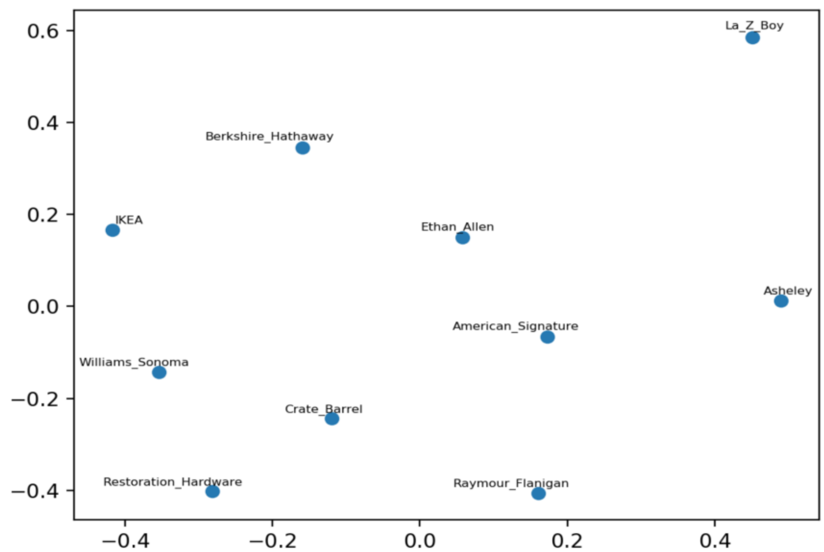


Figure 5: Word Clouds for Image Clusters

Figure 6: Word Clouds for User Clusters

Figure 7: Relative Brand Positions



Online Appendix

TA.1 Benchmark Methods

Since there are thousands of images, each with limited observed characteristics, standard choice models using images as products would not work well and would not serve as a credible benchmark. Instead, we compare the predictive performance against four other popular and representative benchmark methods, ranging from social network-based techniques to various models used in the recommender systems. The first benchmark we compare to is the traditional *collaborative filtering* method, which we denote as *CF*. Collaborative filtering is one of the most popular methods used in the industry, particularly in recommender systems. There are many variations of the *CF* method such as memory-based, model-based, user-based, item-based, and content-based. In our context, the number of users is much higher than the number of "items" (images), so we use the item-based collaborative filtering method as the benchmark. Using this method, we first construct a user-image interaction matrix, *M*, with each cell value representing the user's preference on an image: 1 if the user pinned that image, 0 otherwise. Then we directly apply item-based *CF* by computing item-item similarity (vector cosine) based on the curation history of the users. The probability that an image will be pinned by a user is then calculated as a weighted average: $P_{u,i} =$

$\frac{\sum_{all\ similar\ items\ j}(S_{i,j} * R_{u,j})}{\sum_{all\ similar\ items, j} S_{i,j}}$, where $P_{u,i}$ is the predicted probability that user *u* will pin image *i*, $S_{i,j}$ is the similarity score between image *i* and image j, and $R_{u,j}$ is the "score" given by user *u* on item *j,* which in our context is 1 if user *u* pinned image *i* and 0 otherwise. As is done for our proposed approach, here we also select the top *N* images to recommend based on the predicted probability.

Standard collaborative filtering methods operate directly on a user-item preference matrix that is typically highly sparse, which hinders their performance (Miha, et al. 2005). More sophisticated methods have been developed to improve information extract from the matrix. Our second benchmark, *matrix factorization*, denoted as *MF*, is one such method. Matrix factorization is commonly used to learn latent low-dimensional factors by decomposing the matrix of user-item interactions. Factorization algorithms such as singular-value decomposition and nonnegative matrix factorization have spurred many applications in recommender systems. For our *MF* benchmark method, we use the singular-value decomposition approach to decompose the user-image interaction matrix *M* (the same as used in the *CF* benchmark) into a lower rank approximation: $M = U\Sigma V^T$, where *U* conceptually represents how much each user likes an underlying dimension, V^T conceptually represents how relevant each underlying dimension is to each image, and Σ is a diagonal matrix of singular values, which are essentially weights. For recommending images, we approximate the original matrix through *U*, Σ, and V^T. We then recommend the images with the highest predicted preferences that the user has not already pinned.

Information other than the direct interactions between users and images can also be leveraged. For example, consumers who are connected in social networks often have similar interests, and social networks have been leveraged to improve adoption prediction (Hill et al. 2006). Thus, the third benchmark method is the *social network collaborative filtering* method, which we denote as *SNCF*. This method recommends a set of products to a user that have been adopted by her friends. We implement this method in our social curation context as follows: If a user u_i has a list of friends $F = \{u_1, u_2, \ldots, u_A\}$, and as long as one of them pinned the image I_j in the training period, we predict that I_j will be pinned by user u_i in the holdout period.

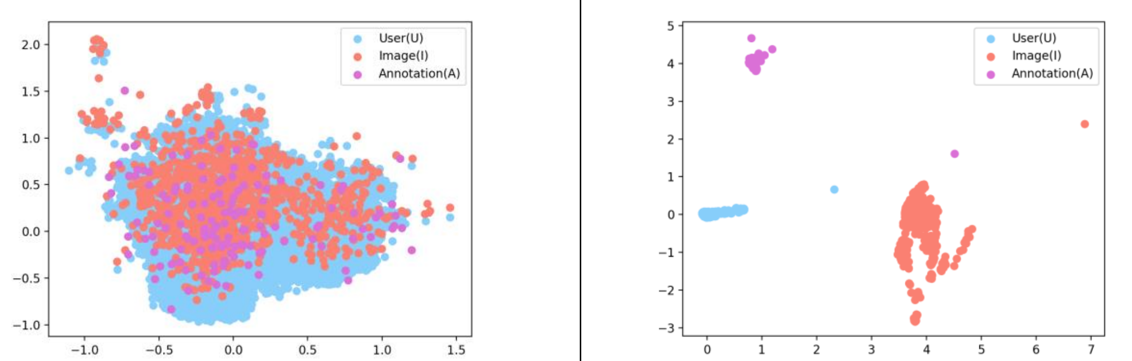
The fourth benchmark we compare our method with is a *community detection-based* method, which we denote as *CD*. Using this method, we first identify distinct groups of images by applying a well-known community detection algorithm called modularity maximization (Newman 2006) on the undirected and non-weighted image-image network where each node is an image and a link is formed if two images have been pinned by the same user in the training period. The algorithm is typically formulated as finding a partition *C* = {C_1, C_2, …, C_k} of the network, where $\forall i, j, C_i \cap C_j = \emptyset$, *k* is the number of communities. The quality of the partition is measured using modularity, which is the fraction of the edges that fall within the given groups minus the expected fraction if edges were distributed at random. Higher modularity values indicate communities with a higher number of intra-community links compared to inter-community links. The modularity is calculated as $Q = \sum_{i=1}^{k}(e_{ii} - a_i^2)$, where $e_{ii} = |\{(u,v): u \in V_i, v \in V_i, (u,v) \in E|/|E|$ is the fraction of edges in community *i*, and $a_i = |\{(u,v): u \in V_i, (u,v) \in E|/|E|$ is the fraction of edges with at least one end in community *i*. We then make recommendations based on the assumption that a focal user u_i will pin an image I_j in the holdout period, as long as the image I_j is in the same community as images that have been pinned by u_i in the training period.

These four benchmark methods represent the state-of-the-art approaches in recommender systems and social network research. The first two benchmarks, *CF* and *MF*, do not model social curation as a network, while the third benchmark, *SNCF*, is based on the follower network of users instead of the heterogeneous information network that we propose. Comparing our proposed approach with these three methods will help establish the merit of our heterogeneous network representation. The last benchmark, *CD*, uses a network representation of the images, but does not use the embedding method which preserves the network's structural and semantic information. Comparing our proposed approach with this last benchmark will help confirm the importance of the network embedding method.

TA.2 Visualization of Raw Embedding Vectors

The network embedding process maps each node in the network to a vector, a 128-dimension one in our study. Since vectors exceeding three dimensions are not amenable to direct visualization, we use principal component analysis (PCA) to extract the first two principal components and generate a 2-D plot of the embedding.

Figure TA.2.1: Visualization of Raw Embedding Vectors



(a) metapath2vec (b) metapath2vec++

Figure TA.2.1 (a) shows the 2-D plot of the embedding vectors generated by *metapath2vec*, and Figure TA.2.1 (b) by *metapath2vec*++. Each dot in the plot corresponds to a user, image, or annotation word, indicated in color and shape. We can see that *metapath2vec*++ groups the nodes of the same type more tightly than *metapath2vec*, and generates larger separation across types. This effect can be understood from Equation (6) in the main paper. As discussed there, *metapath2vec* draws the negative sample from all nodes, while *metapath2vec*++ draws the negative sample only from the nodes of the same type as the context node of the focal node. Optimizing the probability in Equation (6) will push the embedding vector of a negative sample node to be more distant from that of the focal node. By not considering the nodes of the same type as the focal node for a negative sample, *metapath2vec*++ is likely to result in tighter clustering of the same type of nodes and more separation by type, whereas *metapath2vec* will likely not see such a separation by type, since negative samples are drawn without consideration of node type. This distinction provides managers a choice based on their objectives. If their objective calls for intermingled nodes by type, (e.g., to assess the relative distance across node types), then *metapath2vec* is a better fit. However, if they prefer more separation by type (e.g., to create clusters separately for each node type), then *metapath2vec*++ may be better. If the objective is solely to predict future curation actions, then the two can be compared on predictive performance, as shown in section 5.1 of the main paper.

TA.3 Additional Data and Result Tables

Table TA.3.1: Top-20 Tag Words and Board Names

Tag	# Images	Board Name	# Images
patio	64	home	620
ashley	58	idea	592
lamin	50	decor	533
scandinavian	47	project	421
decoupage	38	design	392
outdoor	38	room	366
draw	37	stuff	348
recycle	35	thing	321
refinish	34	craft	314
restor	32	live	309
upcycle	32	interior	295
old	31	dream	283
redo	30	love	281
logo	26	wood	233
showroom	25	paint	225
farmhouse	25	sweet	225
store	25	space	224
metal	24	make	221
vintage	24	bedroom	205
cheap	24	style	194

Table TA.3.2: Average Pins by User-Image Clusters (Holdout Period)

		Image Cluster							
		1	2	3	4	5	6	7	8
User Cluster	1	0.00	0.00	0.00	0.46	0.08	0.00	0.00	0.29
	2	0.03	0.00	0.00	0.78	0.03	0.13	0.00	0.29
	3	0.06	0.00	0.00	0.34	0.06	0.16	0.00	0.46
	4	0.17	0.00	0.00	0.36	0.04	0.07	0.00	0.76
	5	0.16	0.00	0.03	0.34	0.13	0.05	0.00	0.75
	6	0.00	0.00	0.03	2.30	0.13	0.07	0.00	0.56
	7	0.00	0.00	0.00	0.73	0.05	0.08	0.00	0.36
	8	0.14	0.00	0.00	0.37	0.19	0.12	0.00	1.03
	9	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.19
	10	0.36	0.00	0.00	0.60	0.10	0.38	0.00	1.26
	11	0.02	0.00	0.00	0.20	0.38	0.09	0.00	0.36

12	0.10	0.00	0.10	0.26	0.05	0.17	0.00	0.86
13	0.00	0.00	0.02	0.78	2.38	0.05	0.00	0.60
14	0.34	0.00	0.00	0.50	0.12	0.23	0.00	0.44
15	0.00	0.00	0.00	0.37	0.00	2.25	0.00	1.12

All numbers scaled up by 10^4

Correlation between training and holdout sample: 0.66

Table TA.3.3: Annotation Words Closest to Image Clusters

Image Cluster	10 Closest Annotation Words									
1	decoupag	idea	decor	project	craft	stuff	room	thing	design	make
2	ashley	showroom	bedroom	room	decor	great	idea	live	dream	sofa
3	logo	lamin	design	idea	victorian	sketch	thing	craft	décor	project
4	idea	decor	project	design	stuff	garden	outdoor	backyard	patio	thing
5	design	idea	decor	interior	project	stuff	thing	dream	love	room
6	decor	draw	idea	store	project	stuff	craft	dream	design	thing
7	scandinaviar	modular	showroom	ashley	cardboard	design	room	stuff	urban	idea
8	Idea	decor	project	design	room	stuff	thing	craft	interior	dream

Table TA.3.4: Annotation Words Closest to User Clusters

User Cluster	10 Closest Annotation Words									
1	idea	decor	design	project	stuff	dream	thing	craft	interior	live
2	idea	decor	design	project	stuff	thing	room	live	dream	backyard
3	idea	decor	project	design	stuff	room	thing	craft	dream	interior
4	idea	decor	design	stuff	project	thing	room	interior	love	dream
5	idea	decor	design	stuff	room	project	thing	interior	dream	live
6	idea	decor	design	stuff	project	thing	craft	room	love	live
7	idea	decor	design	project	stuff	thing	room	interior	love	dream
8	decor	idea	design	room	project	stuff	thing	interior	dream	live
9	decor	idea	stuff	project	design	thing	room	interior	dream	space
10	idea	decor	project	design	stuff	room	thing	craft	love	interior
11	decor	idea	design	stuff	room	project	thing	interior	dream	love
12	idea	decor	design	thing	project	stuff	room	love	dream	interior
13	idea	design	decor	project	stuff	thing	interior	room	love	dream
14	idea	decor	stuff	project	thing	craft	design	room	dream	love
15	idea	decor	project	stuff	thing	room	craft	interior	love	design

Table TA.3.5: Average Distance between User Clusters and Images by Brand

User Cluster	Williams-Sonoma	American Signature	Ashley	Berkshire Hathaway	Crate & Barrel	Ethan Allen	IKEA	La-Z-Boy	Raymour Flanigan	Restoration Hardware
1	3.724	3.785	3.830	3.610	3.440	3.609	3.975	3.923	3.655	3.452
2	3.666	3.719	3.787	3.536	3.399	3.548	3.896	3.799	3.567	3.444
3	3.718	3.732	3.789	3.581	3.446	3.568	3.931	3.829	3.614	3.445
4	3.755	3.778	3.810	3.585	3.491	3.598	3.984	3.889	3.667	3.477
5	3.706	3.708	3.766	3.552	3.445	3.536	3.933	3.781	3.587	3.454
6	3.709	3.775	3.845	3.587	3.455	3.598	3.931	3.863	3.618	3.488
7	3.686	3.742	3.809	3.568	3.413	3.555	3.922	3.833	3.579	3.449
8	3.730	3.760	3.790	3.558	3.479	3.548	3.972	3.840	3.631	3.508
9	3.791	3.813	3.876	3.641	3.529	3.638	4.011	3.907	3.702	3.571
10	3.694	3.723	3.783	3.580	3.435	3.563	3.930	3.813	3.592	3.389
11	3.771	3.778	3.814	3.601	3.506	3.571	3.981	3.843	3.652	3.556
12	3.757	3.745	3.821	3.569	3.476	3.559	3.973	3.870	3.629	3.476
13	3.771	3.812	3.886	3.667	3.554	3.644	3.967	3.875	3.643	3.571
14	3.783	3.794	3.880	3.699	3.538	3.664	3.988	3.887	3.669	3.536
15	3.668	3.747	3.804	3.564	3.430	3.549	3.949	3.854	3.652	3.424