Network-DPPs: Exploiting User's Network for Interest Prediction in Linkedin Slideshare

Authors
Affiliations
Emails

Abstract

LinkedIn SlideShare is a sharing platform for business documents, videos and presentations which serves as a social discovery platform for users to find relevant content and connect with other members who share similar interests. As new users join the platform, it becomes important to show relevant slide decks which pique user interests in order to sustain user engagement. We wish to leverage insights from user's social network on LinkedIn to better predict users' interests on SlideShare. Determinantal point process (DPP) have recently been shown to be useful for modeling the combinatorial problem of subset selection. In this work, we develop Network-DPPs: Network-aware Determinantal Point Processes which are not only capable of selecting diverse subset of users from a network but also are well equipped to identify key influencing users (*influencers*) whose inclusion would help derive better recommendation. Based on Nystrom low-rank approximations, we derive algorithms for efficient normalizing and sampling from Network-DPPs and demonstrate the usefulness of the model on real world users from LinkedIn Slideshare. Our findings impact the design of interest prediction & recommendation algorithms for better engagement.

1 Introduction

– IN SS
– cold start
– network based personalization
– motivate the need for selecting diverse subset
– motivate the need for considering influencers
– Network-DPPs
– Contributions

LinkedIn has 400M+ members with 100M+ members uniquely visiting per month[cite]. A large fraction of these users also actively use Slideshare for uploading content and satisfying their information need. Started primarily as a business content sharing platform, Slideshare has 18 million presentations, with 400,000 slide decks, on diverse topics ranging from entrepreneurship to music, being uploaded every day. Social signals in the form of likes, shares and downloads are indicators of the popularity of a slideshow and the level of user engagement on topics pertaining to that slideshow.

The user cold-start problem concerns the task of recommending items to users who have not previously purchased or otherwise expressed meaningful preferences towards any items under consideration for recommendation. Addressing the cold-start problem can be important for first-time user engagement and retention and is therefore of critical significance in settings such as online retail.

Over 80% of SlideShare's 70 million unique monthly visitors come through targeted search.[cite] Being primarily a content platform lacking personalized feed, there is little effort by users in building their connections on Slideshare leading to a fairly sparse network. We deal with the sparsity in a novel way by exploiting the property of uniqueness in user's behavioral patterns over multiple social media sites. We leverage the users' LinkedIn network, which is much denser than the Slideshare network, and filter a subset network of Slideshare users.

In essence, we wish to capture the richness of both the platforms, content on Slideshare and connections on LinkedIn.

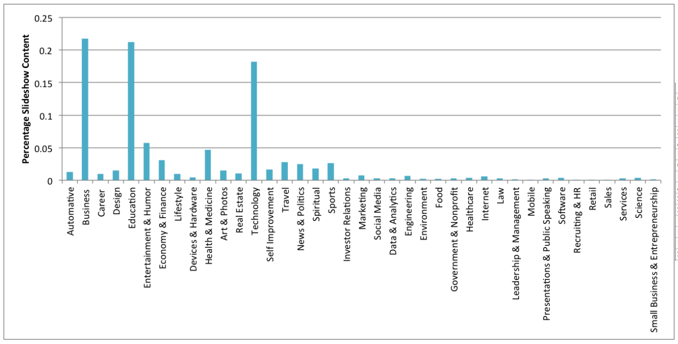
propose a network dependent kernel for DPP

Network based personalization usually takes into account the entire network of a user and it is based on the underlying assumption that all the connections of a user are equally relevant for the user. This assumption might not be true. For some users, only a subset of users might generate relevant leads, for others, almost the entire network can be representative.

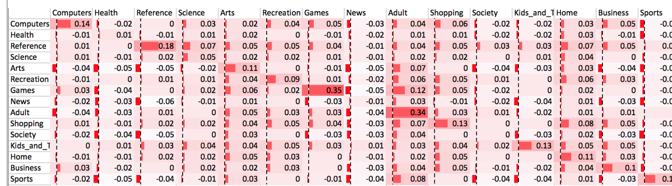
2 Problem Formulation

With more than 10 million presentation uploads, drawing 50 million visitors and 3 billion views a month, Slideshare is a rapidly growing platform with new users signing up every second. In order to engage new users, it becomes important to show relevant presentations that pique their interest. Thus, predicting user's interests becomes an important problem when deriving recommendations. Our main goal in this work is to make better predictions about user's interests from their network information.

Each slideshow on Slideshare is tagged by one or more 'topics' picked from a pool of 39 pre-defined interests - like Real Estate, Health & Medicince, Spiritual, etc. Each



(a) A subfigure



(b) A subfigure

Figure 1: A figure with two subfigures

Symbol	**Description**
n_T	number of children of tree T
$leaves(T)$	set of queries at the leaves of tree T
$ab\|c$	partition of set $\{a, b, c\}$ into disjoint sets $\{a, b\}$,$\{c\}$
ch(T)	children of T
$\phi(T)$	partition of tree T
$\Phi(T)$	set of all partitions ϕ consistent with tree T
$p(\phi(T))$	mixing proportions of partition $\phi(T)$
$\mathbb{H}(T)$	set of all partitions of queries $Q = leaves(T)$
$f(Q)$	task affinity function for set of queries Q

Table 1: Table of symbols

slideshow is tagged with a number of topics, out of which exactly one is selected by the uploader of the slideshow, and the rest are discovered by a topic model. Figure 1 shows an example of [TODO]. Given a user U, we consider the set of users belonging to U's network (N_U) (denoted as "*connections*"). In order to make predictions about users' interests, we make use of their connections. While we could make use of the entire network for making predictions, but as we show later, using all the connections results in low precision scores. Instead, we wish to find the most representative subset of connections from a given user's network which best predict the given user's interests. We explore a number of features which help select the most appropriate subset, including users reputation scores. This falls in line with our intuition that a more reputed user would help us make better recommendations.

3 Network-DPP

3.1 Background/Preliminaries

A point process P on a discrete set $Y = \{1, ..., N\}$ (for example, a collection of documents or images) is a probability measure on 2^Y , the set of all subsets of Y. P is called a determinantal point process (DPP) if, when Y is a random set drawn according to P, we have, for every $A \subseteq Y$:

$$P(A \subseteq Y) = det(K_A) \tag{1}$$

for some positive semidefinite matrix $K \preceq I$ (all eigenvalues of K are less then or equal to 1) indexed by the elements of Y. $K_A \equiv [K_{ij}]_{i,j \in A}$ denotes the restriction of K to the entries indexed by elements of A, and we adopt $det(K_\emptyset) = 1$. We will refer to K as the marginal kernel, as it contains all the information needed to compute the probability of any subset A being included in Y . A few simple observations follow from the equation mentioned above:

$$P(i \in Y) = K_{ii} \tag{2}$$

$$P(i, j \in Y) = K_{ii}K_{jj}K_{ij}K_{ji} = P(i \in Y)P(j \in Y)K_{ij}^2 \tag{3}$$

That is, the diagonal of K gives the marginal probabilities of inclusion for individual elements of Y, and the off-diagonal elements determine the (anti-) correlations between pairs of elements: large values of K_{ij} imply that i and j tend not to co-occur. A DPP might therefore be used naturally to model diverse sets of items, for example in response to a search query. Note that DPPs cannot represent distributions where elements are more likely to co-occur than if they were independent; correlations are always negative. Figure 1 shows the difference between sampling a set of points in the plane using a DPP (with K_{ij} inversely related to the distance between points i and j), which leads to a widely spread set with good coverage, and sampling points independently, where the points exhibit random clumping. Determinantal point processes, introduced to model fermions (Macchi, 1975), also arise in studies of non-intersecting random paths, random spanning trees, and eigenvalues of random matrices (Daley & Vere-Jones, 2003; Borodin & Soshnikov, 2003; Hough et al., 2006).

For the purposes of modeling real data, however, the most relevant construction of DPPs is not through K but via L-ensembles (Borodin, 2009). An L-ensemble defines a DPP via a positive semidefinite matrix L indexed by the elements of Y:

$$P_L(\mathbf{Y} = Y) = \frac{det(L_Y)}{det(L + I)} \tag{4}$$

where I is the N N identity matrix. As a shorthand, we will write $P_L(Y)$ instead of $P_L(\mathbf{Y} = Y)$ when the meaning is clear. Note that PL is normalized due to the identity

$$\sum_{Y \in Y} det(L_Y) = det(L + I) \tag{5}$$

K and L offer alternative representations of DPPs, and we can easily translate between the two; for example, we can compute the marginal kernel K for an L-ensemble:

$$K = (L + I)^1 L \tag{6}$$

Note that K can be computed from an eigendecomposition of $L = \sum_{n=1}^{N} \lambda_n v_n v_n^T$ by a simple rescaling of eigenvalues:

$$K = \sum_{n=1}^{N} \frac{\lambda_n}{\lambda_n + 1} v_n v_n^T \tag{7}$$

We can also similarly compute $L = K(IK)^1$, as long as the inverse exists.

Under both K and L representations, subsets that have higher diversity, as measured by the corresponding kernel, have higher likelihood. However, while K gives rise to marginal probabilities, L-ensembles directly model the probabilities of (exactly) observing each subset of Y, which offers a convenient target for optimization. Furthermore, L need only be positive semidefinite, while the eigenvalues of K are bounded above. For these reasons we focus our modeling efforts on DPPs represented as L-ensembles.

Dual Representation

In special cases where L is a linear kernel of low dimension, Kulesza and Taskar (2010) showed that the complexity of sampling from these DPPs can be be significantly reduced. Let B be the D N matrix whose columns are given by $Bi = q_i \phi_i$, so that $L = B^T B$. Considering the dual kernel matrix $C = B^T B$, L and C share the same nonzero eigenvalues, and for each eigenvector v_k of L,Bv_k is the corresponding eigenvector of C. This leads to the sampling algorithm given in Algorithm 2, which takes time O(D^3 + ND) and space O(ND).

3.2 User Network as DPP

Given the network connections (N_U) of a user U, our goal is to find a representative subset of users from the network connection. Based on the dual representation described above, we cast the network based user subset selection problem in the DPP framework. More specifically, let $q_i \in \mathbb{R}^+$ denote the intrinsic quality of the user i and $\phi_i, \phi_j \in \mathbb{R}^n$ denote the unit length feature vectors representing the similarity between users i and j with $\phi_i^T \phi_j \in [-1, 1]$. Under this framework, we can model quality and similarity separately to encourage the DPP to choose high quality items that are dissimilar to each other.

User's Quality

: When recommending slide decks, a recommendation from a trusted (*reputed*) user values more than recommendation from any randomly selected user. When selecting a subset of connections from a user's network to personalize content, the algorithm should try to select higher quality users (i.e. highly reputed users) so as to obtain higher quality recommendations. We compute user reputation score by taking into account all the social features - views, likes, shares, downloads and embeds - over all the content a user has uploaded on Slideshare. The score over various social features is calculated using a squash function

$$q_i = \tag{8}$$

where c_i is the pre-calculated constant per social signal. Uploading of spammy or near duplicate content results in penalization of User Reputation score.

Quantifying Similarity

: Recall from Section 2 that a user is represented as a distribution over the interest profile. Because a user's probabilistic profile is a distribution, it enables a principled comparison between any two users, by comparing their interest profile distributions using information theoretic divergence metrics. In order to quantify the preferential similarity between the user and his connection, we compute the Jensen-Shannon (JS) Divergence [Dagan *et al.*, 1997] to compare the similarity or distance between the full probability distributions of two users:

$$\phi_i^T \phi_j = JS(u_i||u_j) = KL(u_i||\frac{u_i + u_j}{2}) + KL(u_j||\frac{u_i + u_j}{2}) \tag{9}$$

where KL is the KL-divergence score between user profiles u_i and u_j. To handle the zero frequency problem in calculating KL divergence, we used absolute discounting with ϵ = 0.001. TODO: convert JSD to similarity metric by normalizing and subtracting from 1.

With the quality and similarity function defined as above, we can construct the dual kernel matrix C as follows:

$$C = B^T B = q_i \phi_i^T \phi_j q_j \tag{10}$$

The dual kernel matrix can then

3.3 Sampling

Algorithm 1, due to TODO: NIPS 2010, gives an efficient algorithm for sampling a configuration Y from a DPP. The input to the algorithm is an eigendecomposition of the dual kernel matrix C. Algorithm 1 has two main loops, corresponding to two phases of sampling. In the first phase, a subset of the eigenvectors is selected at random, where the probability of selecting each eigenvector depends on its associated eigenvalue. In the second phase, a sample Y is produced based on the selected vectors. Note that on each iteration of the second loop, the cardinality of Y increases by one and the dimension of V is reduced by one. The sampled configuration Y denotes the subset of users sampled form the network which are to be used for predicting user's Slideshare interest. This configuration of the network DPP is the vanilla version which we refer to as vanilla DPP and use as a baseline to compare against in our experiments.

The vanilla-DPP suffers from various issues including infeasible sampling as network grows large as well as onadaptivity to the amount of data per user. We next describe a data-adaptive version of the vanilla DPP that addresses these issues and present a modified sampling approach for the same.

4 Influencer based Nystrom Approximation

– large network – adaptive BNP – informative columns

While users' first degree connections are relatively small in size, the network size exponentially increases when we consider 2^{nd}-degree connections to form the network. As the network size grows, sampling from DPP becomes computationally infeasible. To combat the increased network size, we consider applying the Nystrom approximation to project the kernel matrix into a low-dimensional space [Affandi *et al.*, 2013]. The Nystrom method is an efficient technique to generate low-rank matrix approximations and is used in several large-scale learning applications. Given a kernel matrix K, the Nystrom method can be deemed as choosing a subset of m columns (*landmarks*) $E \in R^{nm}$, and reconstructing the complete kernel matrix by $K \cong EW^1E^T$, where W is the intersection of the selected rows and columns of K.

A key aspect of this method is the distribution according to which landmarks (columns) are sampled from the original matrix and then used as the basis for a low rank approximation. The most popular sampling scheme for Nystrom method is random sampling for a fixed number of landmarks, which leads to fast versions of kernel machines. However, since different users have different network sizes, such a naive Nystrom sampling technique with fixed landmark size cannot be used. This motivates the need for an adaptive landmark selection scheme based on the user's network which takes into account the network in consideration while deciding on the landmarks and their number.

4.1 Adaptive Landmark Selection

We wish to leverage a user's network information while deciding on the landmarks to be used for the Nystrom approximation. A large connection network implies a large kernel matrix and as a result, necessitates selecting a larger number of landmarks. Additionally, when deciding on the landmarks, we wish to capture insights on which set of connections to include in the landmark, so as to obtain best performance. In order to accomplish these tasks, we propose the use of nonparametric clustering of the user's network into network clusters and selecting representative landmarks from each of the obtained clusters. We next describe a bayesian nonparametric model which achieves the same.

Network Clustering with Nonparametric Priors

The Chinese restaurant process (CRP) is a distribution on all possible partitions of a set of objects (in our case, network connections). The generative process can be described via a restaurant with an infinite number of tables (in our case, clusters). Customers (connections) i enter the restaurant in sequence and select a table z_i to join. They pick an occupied table with a probability proportional to the number of customers already sitting there, or a new table with probability proportional to a scaling parameter α. The dd-CRP alters the CRP by modeling customer links not to tables, but to other customers.

In our network clustering problem, each network is associated with a dd-CRP and its tables are embellished with IID draws from a base distribution over mixture component parameters. Let c_i denote the ith connection assignment, the index of the query with whom the ith connection is linked. Let d_{ij} denote the distance measurement between connections i and j, let D denote the set of all distance measurements between connections, and let f be a decay function. The distance dependent CRP independently draws the connection assignments to clusters conditioned on the distance measurements,

$$p(c_i = j \| D, \alpha) \propto \begin{cases} f(d_{ij}) & \text{if } j \neq i \\ \alpha & \text{if } j = i \end{cases}$$

Here, d_{ij} is an externally specified distance between connections i and j, and α determines the probability that a customer links to themselves rather than another customer. The monotonically decreasing decay function $f(d)$ mediates how the distance between two queries affects their probability of connecting to each other. Given a decay function f, distances between connections D, scaling parameter α, and an exchangeable Dirichlet distribution with parameter λ, N connections are drawn as follows, [TODO: change the Dirichlet-Mult to the other word level case in the dd-crp journal paper]

1. For $i \in [1, N]$, draw $c_i \sim dist - CRP(, f, D)$.
2. For $i \in [1, N]$,
   (a) If $c_i \notin R^*_{q_{1:N}}$, set the parameter for the ith query to $\theta_i = \theta_{q_i}$. Otherwise draw the parameter from the base distribution, $\theta_i \sim Dirichlet(\lambda)$.
   (b) Draw the ith query, $w_i \sim Mult(M, \theta_i)$.

We make use of the Jensen-Shannon (JS) Divergence between connections (described in Section 3.2) to define the distance function between connections. The posterior of the proposed dd-CRP model is intractable to compute because the dd-CRP places a prior over a combinatorial number of possible customer configurations. We employ Gibbs sampling wherein we iteratively draw from the conditional distribution of each latent variable given the other latent variables and observations. The Gibbs sampler iteratively draws from

$$\begin{aligned} p(q_i^{new}|q_{-i}, x) \propto\ & p(q_i^{new})|D, \alpha) \\ & p(x|z(c_{-i} \cup c_i^{new}), G_0) \end{aligned} \tag{11}$$

The first term is the dd-CRP prior and the second term is the likelihood of the observations under the partition.

Cluster specific Landmarks

With the clusters identified, we propose to select landmark points from each cluster in order to leverage the encoding powers of the landmark points in summarizing the data. Let l be the total number of landmark points to be selected. The larger the l, the more accurate the Nystrom approximation though at the cost of higher computations. Since different users would have different number of connections, fixing l for all users limits the expressive powers of user's connection network. The nonparametric clustering scheme described above finds the number of clusters based on the network data and as result could inform our choice for the network specific l. We propose to select two landmark points per cluster:

1. **Cluster Mean:**
   Clustering based approximations were recently shown ([Zhang *et al.*, 2008]) to improve the approximation quality compared to other sampling schemes. From among the set of connections defining the cluster, we

select the connection who has the lowest mean distance (JS Divergence) to all other connections in the cluster.

2. **Most Reputed User:**
In addition to the cluster mean, we also select the most reputed user per cluster in an attempt to include high quality users among the set of landmarks. We select reputed users in an attempt to pick columns which are informative in terms of interest predictions and recommendations.

4.2 Nystorm Approximation

Given a sample W of l landmark items corresponding to a subset of the indices of an N N symmetric positive semidefinite matrix L, let $\overline{W}$ be the complement of W (with size Nl), let L_W and $L_{\overline{W}}$ denote the principal submatrices indexed by W and $\overline{W}$, respectively, and let $L_{\overline{W}W}$ denote the $(Nl) \times l$ submatrix of L with row indices from $\overline{W}$ and column indices from W. Then we can write L in block form as

$$L = \begin{pmatrix} L_W & L_{W\overline{W}} \\ L_{\overline{W}W} & L_{\overline{W}} \end{pmatrix}$$

If we denote the pseudo-inverse of L_W as L_W^+ , then the Nystrom approximation of L using W is

$$\widetilde{L} = \begin{pmatrix} L_W & L_{W\overline{W}} \\ L_{\overline{W}W} & L_{W\overline{W}} L_W^+ L_{\overline{W}W} \end{pmatrix}$$

Fundamental to this method is the choice of W for which we employed a clustered model as already described in Section 4.1.

4.3 Sampling

Given the network kernel matrix L, we propose applying the Nystrom approximation to L, building an l-dimensional approximation as above and applying the dual representation sampling to the approximated matrix. Algorithm 2 described the steps in obtaining the subset of connections from a user's connection network.

5 Experimental Evaluation

The main goal of our work is to most informatively select connections from a user's LinkedIn connection network which are most representative of the user's Slideshare interests. We evaluate the proposed connection subset selection model on the task of interest prediction. As described in Section 2, we construct an interest profile from user's Slideshare likes and evaluate how well the selected subset of connections is able to predict the user's interest. The user's probability distribution vector is sorted to extract top 'k' topics, which act as the ranking gold-standard. We quantify the quality and coverage of the inferred interests based on the a number of metrics including Precision, Recall, F1 score and Average Precision.

5.1 Data

We consider real world data of over 29 million users with over 18 million presentations, with 400,000 slide decks, on diverse topics ranging from entrepreneurship to music. In order to gauge a user's topical interests, we take user's likes on slideshows into consideration. We consider a random sample of users from Slideshare which follow the requirement of having atleast one like on slideshare, and number of LinkedIn connections (active on slideshare) greater than the threshold of 250. The average number of connections per user for the users in purview of our analysis is 448. Note that to identify interests of new users on Slideshare, we leverage his existing LinkedIn connections who might already be engaged on Slideshare. Note that the users in this underlying network may not be connected on Slideshare.

For each user, a distribution of his level of interest over the 39 topics is drawn, by considering the slideshows he has shown interest in. A 39X1 stochastic vector, representing the topical probability distribution, is created per user by taking into account the topics pertaining to the user's liked slideshows. In order to quantify the preferential similarity between the user and his connections, we compute the Jenson-Shennon divergence score between the probability distribution of user and his connections.

5.2 Baselines

We consider a number of baselines to compare our connection selection framework:

- **Most Similar Users (Top 10%):**
Our intuition was that personalization based on the most similar connections in a user's network would lead to highly relevant results, but it might result in a loss of coverage across the topics the user is interested in. Motivated by this intuition, we extracted top 10% similar connections of the users, using the Jenson-Shannon divergence scores between the topical interest profiles of uses, and use these most similar connections in making predictions about the interests for the given user.
- **Vanilla DDP:**
As described in Section 3.2, the vanilla DPP baseline selects the subset based on the quality vs diversity tradeoff without giving importance to reputed users.
- **Random 10%:**
In this baseline, we randomly select a subset of users from the connection network.
- **Entire Network:**
To represent the probability distribution over the topics for the entire network, we take the centroid of all connection vectors, and use it to predict the top interests for the user.

5.3 Results

6 Related Work

We cover several areas of related work and discuss how our work relates to and extends prior work.

Network based personalization

With the emergence of social networks, recommender systems that rely on SNS have started to gain popularity. For example, Said et al. [Said *et al.*, 2010] explored movie recommendations based on user ratings and a social network (that comes with the user ratings data) where users can befriend one another. Jamali and Ester [Jamali and Ester, 2010] incorporated trust propagation mechanisms into the matrix factorization technique in order to recommend items. In addition, given an external social network, a standard way of profiling users is through the identification of influence [Bakshy *et al.*, 2011], which can be measured in a number of ways. Kwak et al. [Kwak *et al.*, 2010] compared three different measures of influence on Twitter (number of followers, PageRank [Page *et al.*, 1999], and number of retweets) and found that the ranking of the most influential users differed depending on the measure.

Pham et al. [Pham *et al.*, 2011] find user and item clusters in social networks and use such information to enhance CF methods. Random walk has also been exploited for recommendation. Yildirim and Krishnamoorthy [Yildirim and Krishnamoorthy, 2008] build a graph of items, in which each link is weighed by the similarity of its two owner items. Konstas et al. [Konstas *et al.*, 2009] study CF methods on a music social network to predict music playcounts.

Cold start recommendations

As the lack of ratings (i.e., the cold-start) hinders the use of CF techniques [Schein *et al.*, 2002], various alternatives have been employed to overcome the problem. For example, Zhou et al. [Zhou *et al.*, 2011] experimented with eliciting new user preferences by using decision trees with collaborative filtering. Moshfeghi et al. [Moshfeghi *et al.*, 2011] proposed a method for combining content features such as semantic and emotion information with ratings information for the recommendation task. Liu et al. [Liu *et al.*, 2011] identified representative users whose linear combinations of tastes are able to approximate other users.

Deteminantal Point Processes

DPPs have been studied in statistical physics and probability [Borodin, 2009; Borodin and Rains, 2005] and have recently gained substantial interest in machine learning [Gillenwater *et al.*, 2012; **?**; **?**; **?**]. Gillen et al. [Gillenwater *et al.*, 2012] model document threads as DPPs and use it to summarize document collections. Gong et al. [Gong *et al.*, 2014] model video sequences as DPPs and perform supervised video summarization. DPPs have also been used as diversity priors in generative models [Kwok and Adams, 2012].

7 Conclusion

References

[Affandi *et al.*, 2013] Raja Hafiz Affandi, Alex Kulesza, Emily Fox, and Ben Taskar. Nystrom approximation for large-scale determinantal processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pages 85–98, 2013.

[Bakshy *et al.*, 2011] Eytan Bakshy, Jake M Hofman, Winter A Mason, and Duncan J Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 65–74. ACM, 2011.

[Borodin and Rains, 2005] Alexei Borodin and Eric M Rains. Eynard–mehta theorem, schur process, and their pfaffian analogs. *Journal of statistical physics*, 121(3-4):291–317, 2005.

[Borodin, 2009] Alexei Borodin. Determinantal point processes. *arXiv preprint arXiv:0911.1153*, 2009.

[Dagan *et al.*, 1997] Ido Dagan, Lillian Lee, and Fernando Pereira. Similarity-based methods for word sense disambiguation. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pages 56–63. Association for Computational Linguistics, 1997.

[Gillenwater *et al.*, 2012] Jennifer Gillenwater, Alex Kulesza, and Ben Taskar. Discovering diverse and salient threads in document collections. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 710–720. Association for Computational Linguistics, 2012.

[Gong *et al.*, 2014] Boqing Gong, Wei-Lun Chao, Kristen Grauman, and Fei Sha. Diverse sequential subset selection for supervised video summarization. In *Advances in Neural Information Processing Systems*, pages 2069–2077, 2014.

[Jamali and Ester, 2010] Mohsen Jamali and Martin Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 135–142. ACM, 2010.

[Konstas *et al.*, 2009] Ioannis Konstas, Vassilios Stathopoulos, and Joemon M Jose. On social networks and collaborative recommendation. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 195–202. ACM, 2009.

[Kwak *et al.*, 2010] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.

[Kwok and Adams, 2012] James T Kwok and Ryan P Adams. Priors for diversity in generative latent variable models. In *Advances in Neural Information Processing Systems*, pages 2996–3004, 2012.

[Liu *et al.*, 2011] Nathan N Liu, Xiangrui Meng, Chao Liu, and Qiang Yang. Wisdom of the better few: cold start recommendation via representative based rating elicitation. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 37–44. ACM, 2011.

[Moshfeghi *et al.*, 2011] Yashar Moshfeghi, Benjamin Piwowarski, and Joemon M Jose. Handling data sparsity in collaborative filtering using emotion and semantic based features. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 625–634. ACM, 2011.

[Page *et al.*, 1999] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: bringing order to the web. 1999.

[Pham *et al.*, 2011] Manh Cuong Pham, Yiwei Cao, Ralf Klamma, and Matthias Jarke. A clustering approach for collaborative filtering recommendation using social network analysis. *J. UCS*, 17(4):583–604, 2011.

[Said *et al.*, 2010] Alan Said, Ernesto W De Luca, and Sahin Albayrak. How social relationships affect user similarities. In *Proc. of the 2010 Workshop on Social Recommender Systems*, pages 1–4, 2010.

[Schein *et al.*, 2002] Andrew I Schein, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 253–260. ACM, 2002.

[Yildirim and Krishnamoorthy, 2008] Hilmi Yildirim and Mukkai S Krishnamoorthy. A random walk method for alleviating the sparsity problem in collaborative filtering. In *Proceedings of the 2008 ACM conference on Recommender systems*, pages 131–138. ACM, 2008.

[Zhang *et al.*, 2008] Kai Zhang, Ivor W Tsang, and James T Kwok. Improved nyström low-rank approximation and error analysis. In *Proceedings of the 25th international conference on Machine learning*, pages 1232–1239. ACM, 2008.

[Zhou *et al.*, 2011] Ke Zhou, Shuang-Hong Yang, and Hongyuan Zha. Functional matrix factorizations for cold-start recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 315–324. ACM, 2011.