Deep geometric matrix completion: Are we doing it right?

Amit Boyarski* **Sanketh Vedula*** **Alex M. Bronstein**

Technion – Israel Institute of Technology, Haifa, Israel

{amitboy, sanketh, bron}@cs.technion.ac.il

Abstract

We address the problem of reconstructing a matrix from a subset of its entries. Current methods, branded as *geometric deep learning*, augment classical rank regularization techniques by incorporating geometric information into the solution. This information is usually provided as graphs encoding relations between rows/columns. In this work we propose a simple spectral approach for solving the matrix completion problem. We introduce the *zoomout loss*, a multiresolution spectral geometric loss inspired by recent advances in shape correspondence, whose minimization leads to state-of-the-art results on various recommender systems datasets. Surprisingly, for some datasets we were able to achieve comparable results even without incorporating geometric information. This puts into question both the quality of such information and current methods' ability to use it in a meaningful and efficient way.

1 Introduction

Matrix completion deals with the recovery of missing values of a matrix from a subset of its entries,

$$\text{Find } \boldsymbol{X} \text{ s.t. } \boldsymbol{X} \odot \boldsymbol{S} = \boldsymbol{M} \odot \boldsymbol{S}. \tag{1}$$

Here $\boldsymbol{X}$ stands for the unknown matrix, $\boldsymbol{M} \in \mathbb{R}^{m \times n}$ for the ground truth matrix, $\boldsymbol{S}$ is a binary mask representing the input support, and $\odot$ denotes the Hadamard product. In what follows, we shall use the nomenclature of the machine learning literature, referring to the given entries of $\boldsymbol{M}$ as *trainings set*. Since problem (1) is ill-posed, it is common to assume that $\boldsymbol{M}$ belongs to some low dimensional subspace. Under this assumption, the matrix completion problem can be cast via the least-squares variant,

$$\min_{\boldsymbol{X}} \operatorname{rank}(\boldsymbol{X}) + \frac{\mu}{2} \|(\boldsymbol{X} - \boldsymbol{M}) \odot \boldsymbol{S}\|_F^2 . \tag{2}$$

Relaxing the intractable rank penalty to its convex envelope, namely the nuclear norm, leads to a convex problem whose solution coincides with (2) under some technical conditions [Candès and Recht, 2009]. Another way to enforce low rank is by explicitly parametrizing $\boldsymbol{X}$ in factorized form, $\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{Z}^\top$. The rank is upper-bounded by the inner dimensions of $\boldsymbol{Y}, \boldsymbol{Z}^\top$. A recent study by Arora et al. [2019] suggests that overparametrizing $\boldsymbol{X}$ as a product of L matrices results in a low rank matrix due to implicit regularization properties of the gradient descent algorithm. This result is surprising since such models are a priori expected to overfit the training data.

Current approaches for matrix completion, falling under the umbrella of *geometric deep learning*, make use of side information in the form of graphs encoding relations between rows/columnsKalofolias et al. [2014], Monti et al. [2017]. Geometric deep learning generalizes the standard deep learning approaches, tailored for structured Euclidean domains, to domains such as general graphs and manifolds. In particular, *graph convolutional neural networks* (GCNNs) follow the standard architecture of Euclidean CNNs, but replace the Euclidean convolution operator with linear filters constructed using the graph Laplacian. While these techniques achieve state-of-the-art results in various tasks, their design is arguably cumbersome and non-intuitive. It has recently been demonstrated that simple architectures perform competitively on several graph analysis tasks [Wu et al., 2019]. Such simple techniques have the advantage of being easier to analyze and reproduce.

*Equal contribution.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

Related work. The literature on matrix completion is vast and it is impossible to fully review it in this brief note. Of particular significance is the pioneering work of Candès and Recht [2009], which initiated line of works on nuclear norm minimization. For a contemporary review of these methods the reader is referred to [Li et al., 2019] and the references therein. Kalofolias et al. [2014] were the first to introduce graphs into the convex low-rank matrix recovery problem. Later approaches integrated these graphs into the deep learning framework in what is now called geometric deep learning. We note the works of Monti et al. [2017] who combined geometric matrix completion with recurrent multi-graph neural networks and Berg et al. [2017] who used a graph autoencoder framework based on differentiable message passing to predict the missing entries. The inspiration for this paper stems from techniques for finding shape correspondence. In particular, the functional maps framework and its variants [Ovsjanikov et al., 2012, 2016], most notably the work of [Litany et al., 2017] who combined functional maps with joint diagonalization to solve partial shape matching problems, and the *product manifold filter* (PMF) [Vestner et al., 2017a,b] and *zoomout* [Melzi et al., 2019] – two greedy algorithms for correspondence refinement by gradual introduction of high frequencies. This last method lent its name to the loss we define in Section 3.

Contribution. Our contributions are as follows: Firstly, we provide a geometric interpretation for *deep matrix factorization* [Arora et al., 2019] via the functional maps framework. Secondly, we introduce the *zoomout loss*, a multiresolution spectral geometric loss inspired by recent advances in shape correspondence. Thirdly, we show that via a simple *shallow and fully linear* network, it is possible to obtain state-of-the-art results on various recommendation systems datasets. And lastly, we demonstrate that in some cases the effect of the geometry is only marginal, and results on par with state-of-the-art can be achieved even without it.

2 Background

Spectral graph theory. Let $G = (V, E, \boldsymbol{\Omega})$ be a (weighted) graph specified by its vertex set V and edge set E, and let $\boldsymbol{\Omega}$ be its adjacency matrix. Given a function $\boldsymbol{x} \in \mathbb{R}^{|V|}$ on the vertices, we define the following quadratic form (also known as *Dirichlet energy*) measuring the variability of the function $\boldsymbol{x}$ on the graph,

$$\boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x} = \sum_{(a,b)\in E} \omega_{a,b} \left(x(a) - x(b)\right)^2 . \tag{3}$$

The matrix $\boldsymbol{L}$ is called the *(combinatorial) graph Laplacian*, and is given by $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{\Omega}$, where $\boldsymbol{D} = \operatorname{diag}(\boldsymbol{\Omega}\mathbf{1})$ is the *degree matrix*. $\boldsymbol{L}$ is symmetric and positive semi-definite and therefore admits a spectral decomposition $\boldsymbol{L} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^\top$. The graph Laplacian is a discrete generalization of the continuous Laplace-Beltrami operator, and therefore has similar properties. One can think of the eigenpairs $(\boldsymbol{\phi}_i, \lambda_i)$ as the graph analogues of "harmonic" and "frequency". A function $\boldsymbol{x} = \sum_{i=1}^{|V|} \alpha_i \boldsymbol{\phi}_i$ on the vertices of the graph whose coefficients α_i are small for large i, demonstrates a "smooth" behaviour on the graph in the sense that the function values on nearby nodes will be similar. A standard approach to promoting such smooth functions on graphs is by using the Dirichlet energy (3) to regularize some loss term. For example, this approach gives rise to the popular bilateral and non-local means filters [Gadde et al., 2013].

Functional maps. Let $G_1 = (V_1, E_1, \boldsymbol{\Omega}_1)$, $G_2 = (V_2, E_2, \boldsymbol{\Omega}_2)$ be two graphs, and let $\boldsymbol{\Phi}, \boldsymbol{\Psi}$, be two orthonormal bases for functions defined on the vertices of these graphs. Given two such functions, $\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{\alpha}$ on G_1 and $\boldsymbol{y} = \boldsymbol{\Psi}\boldsymbol{\beta}$ on G_2, one can define a map $\boldsymbol{C}$ between their representations $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, i.e., $\boldsymbol{\alpha} = \boldsymbol{\Phi}^\top \boldsymbol{x} = \boldsymbol{C}\boldsymbol{\Psi}^\top \boldsymbol{y} = \boldsymbol{C}\boldsymbol{\beta}$. The matrix $\boldsymbol{C}$ represents a linear map between the functional spaces on G_1 and G_2, known as a *functional map*. Let $\boldsymbol{X}$ be a function defined on the product graph $G_1 \times G_2$, then the functional map is given by projecting $\boldsymbol{X}$ onto the corresponding bases $\boldsymbol{\Phi}, \boldsymbol{\Psi}$, $\boldsymbol{C} = \boldsymbol{\Phi}^\top \boldsymbol{X} \boldsymbol{\Psi}$. Using the SVD, one can decompose $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ to interpret $\boldsymbol{C}$ as a mapping between a function on one graph and a function on the other. To get $\boldsymbol{X}$ back from the functional map, one can use $\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^\top$. For computational reasons, it is common to use truncated bases $\boldsymbol{\Phi}, \boldsymbol{\Psi}$, in which case the last equality holds only approximately.

The structure of the functional map depends on the properties of the chosen bases and the functions it maps. A common choice for a basis is the aforementioned Laplacian eigenbasis, building on the assumption that the signals involved are smooth with respect to the graphs. While this is a useful model, it assumes that the given graphs encode the geometry in an adequate way. In real world

problems these graphs are only approximate, constructed from heuristic features associated with the row and the column spaces. Given better graphs, a simpler structure of the functional map emerges. For example, by introducing two orthonormal matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$, one can make the functional map $\boldsymbol{C} = (\boldsymbol{\Phi}\boldsymbol{P})^\top \boldsymbol{X} (\boldsymbol{\Psi}\boldsymbol{Q})$ diagonal. These orthonormal matrices can be thought of as a way of aligning the bases $\boldsymbol{\Phi}, \boldsymbol{\Psi}$, with the principal axes of $\boldsymbol{X}$.

3 Spectral geometric matrix completion

We assume that we are given a set of samples from the unknown matrix $\boldsymbol{M} \in \mathbb{R}^{m\times n}$, encoded by a binary mask $\boldsymbol{S}$, and two graphs $G_\mathrm{r}, G_\mathrm{c}$, encoding relations between the rows and the columns, respectively. Denote the Laplacians of these graphs and their spectral decompositions by $\boldsymbol{L}_\mathrm{r} = \boldsymbol{\Phi}\boldsymbol{\Lambda}_\mathrm{r}\boldsymbol{\Phi}^\top$, $\boldsymbol{L}_\mathrm{c} = \boldsymbol{\Psi}\boldsymbol{\Lambda}_\mathrm{c}\boldsymbol{\Psi}^\top$. Our approach relies on a minimization problem of the form

$$\min_{\boldsymbol{X}} \; E(\boldsymbol{X}) + \mu_r \mathrm{trace}\left(\boldsymbol{X}^\top \boldsymbol{L}_\mathrm{r} \boldsymbol{X}\right) + \mu_c \mathrm{trace}\left(\boldsymbol{X}\boldsymbol{L}_\mathrm{c}\boldsymbol{X}^\top\right), \tag{4}$$

with E denoting the data term that we discuss in the sequel. As mentioned above, the input graphs are typically constructed from a set of heuristically gathered features and therefore may provide a poor representation of the latent geometry. One way to account for this inaccuracy could be to include the features in our optimization. This will induce, through a complicated nonlinear dependence, a different metric (i.e., adjacency matrix) and a different graph Laplacian. This approach lies at the heart of smoothing models based on conditional random fields (CRFs) used in image segmentation (see for example Krähenbühl and Koltun [2011]).

We adopt a different approach by working in the spectral domain. Switching to the spectral domain allows us to modify the metric indirectly by applying orthogonal transformations, $\boldsymbol{P}$ and $\boldsymbol{Q}$, to the bases $\boldsymbol{\Psi}, \boldsymbol{\Phi}$. The purpose of these transformations is to rotate $\boldsymbol{\Psi}, \boldsymbol{\Phi}$, in a way that will simplify the structure of $\boldsymbol{C}$. Since our method relies on the premise that the matrix $\boldsymbol{M}$ is smooth with respect to some graphs, our interest is in modified bases $\boldsymbol{\Phi}\boldsymbol{P}, \boldsymbol{\Psi}\boldsymbol{Q}$, which arise from the eigendecomposition of a graph Laplacian. To that end, we shall use $\boldsymbol{L}_\mathrm{r}, \boldsymbol{L}_\mathrm{c}$, as proxies for the latent graph Laplacians, and promote bases that approximately diagonalize them by introducing two energy terms, $E^r_\mathrm{diag} = \left\|\mathrm{off}\left(\boldsymbol{P}^\top \boldsymbol{\Lambda}_\mathrm{r} \boldsymbol{P}\right)\right\|_F^2$, and $E^c_\mathrm{diag} = \left\|\mathrm{off}\left(\boldsymbol{Q}^\top \boldsymbol{\Lambda}_\mathrm{c} \boldsymbol{Q}\right)\right\|_F^2$, where $\mathrm{off}(\cdot)$ denotes the off-diagonal elements. Under these modifications, the Dirichlet energy terms from (4) become, $E^r_\mathrm{dir} = \mathrm{trace}\left(\boldsymbol{Q}\boldsymbol{C}^\top \boldsymbol{P}^\top \boldsymbol{\Lambda}_\mathrm{r} \boldsymbol{P}\boldsymbol{C}\boldsymbol{Q}^\top\right)$ and $E^c_\mathrm{dir} = \mathrm{trace}\left(\boldsymbol{P}\boldsymbol{C}\boldsymbol{Q}^\top \boldsymbol{\Lambda}_\mathrm{c} \boldsymbol{Q}\boldsymbol{C}^\top \boldsymbol{P}^\top\right)$. Finally, we introduce the following energy terms to promote the (approximate) orthonormality of $\boldsymbol{P}, \boldsymbol{Q}$: $E^r_\mathrm{orth}(\boldsymbol{P}) = \left\|\boldsymbol{P}^\top \boldsymbol{P} - \boldsymbol{I}\right\|_F^2$, $E^c_\mathrm{orth}(\boldsymbol{Q}) = \left\|\boldsymbol{Q}^\top \boldsymbol{Q} - \boldsymbol{I}\right\|_F^2$.

Zoomout loss. Let us denote the training error achieved by a matrix $\boldsymbol{X}_{p,q} \equiv \boldsymbol{\Phi}_p \boldsymbol{C}_{p,q} \boldsymbol{\Psi}_q^\top$ synthesized from the first p vectors in $\boldsymbol{\Phi}$ and the first q vectors in $\boldsymbol{\Psi}$ as $Z_{p,q} \equiv \| \left(\boldsymbol{\Phi}_p \boldsymbol{C}_{p,q} \boldsymbol{\Psi}_q^\top - \boldsymbol{M}\right) \odot \boldsymbol{S}\|_F^2$. We define the zoomout loss as follows:

$$E_\mathrm{z} = \sum_{p=1,q=1}^{m,n} w_{p,q} Z_{p,q}, \tag{5}$$

with the weights $w_{p,q} \geq 0$. It can be shown that it is enough to use each value of p or q only once and therefore most of the $w_{p,q}$ shall be set to 0. We defer those details for an extended version of this article. The zoomout loss (5) is inspired by the greedy approaches for shape correspondence proposed by Vestner et al. [2017b], Melzi et al. [2019], but we advocate minimizing the training error using all resolutions at once. This simultaneous multi-resolution approach incurs a penalty on the rank of the reconstructed matrix by implicitly giving more weight to the low frequency terms. For example, the sub-matrix $\boldsymbol{C}_{2,2}$ appears in all the terms $Z_{p,q}$ with $p \geq 2, q \geq 2$, emphasizing its importance. This allows us to be sloppy in the estimation of the rank of $\boldsymbol{M}$ without compromising the results by much. Following the discussion above, we replace $\boldsymbol{\Phi}, \boldsymbol{\Psi}$ with $\boldsymbol{\Phi}\boldsymbol{P}, \boldsymbol{\Psi}\boldsymbol{Q}$, obtaining $Z_{p,q}(\boldsymbol{P}, \boldsymbol{C}, \boldsymbol{Q}) \equiv \| \left(\boldsymbol{\Phi}\boldsymbol{P}_p \boldsymbol{C}_{p,q} \boldsymbol{Q}_q^\top \boldsymbol{\Psi}^\top - \boldsymbol{M}\right) \odot \boldsymbol{S}\|_F^2$. With this modification, the zoomout loss (5) should favor bases $\boldsymbol{\Phi}\boldsymbol{P}$, $\boldsymbol{\Psi}\boldsymbol{Q}$, in which most of the energy of $\boldsymbol{C}$ is concentrated in the low frequencies, i.e., the top-left part. An interesting observation is that by setting $p = m, q = n$, we get the *deep matrix factorization* (DMF) method from Arora et al. [2019] (up to initialization). As we show in Section 4, this model alone, without any additional geometric priors, is sufficient to obtain results on par with the state-of-the-art on some datasets. This puts in question the quality of the geometric information in those cases. The complete minimization objective combines all the described terms

Model	Synthetic Neflix	Flixster	Douban	ML-100K
MC [Candès and Recht, 2009]	–	1.533	0.845	0.973
GMC [Kalofolias et al., 2014]	0.3693	–	–	0.996
GRALS [Rao et al., 2015]	0.0114	1.313/1.245	0.833	0.945
RGCNN [Monti et al., 2017]	0.0053[a]	1.179/0.926	0.801	0.929
GC-MC [Berg et al., 2017]	–	**0.941**/0.917	0.734	0.910[b]
FM (ours)	0.0064	–	–	–
DMF [Arora et al., 2019], (ours)	0.215	1.06	0.732	0.92[c]/ 0.922
SGMC (ours)	**0.0021**	0.971 / 0.900	**0.731**	0.912
SGMC-Z (ours)	0.0036	0.957 / **0.888**	0.733	**0.907**[c]/ 0.913

[a] This number corresponds to the inseparable version of MGCNN.
[b] This number corresponds to GC-MC.
[c] Early stopping.

Table 1: RMSE test set scores for runs on Synthetic Netflix [Monti et al., 2017], Flixster [Jamali and Ester, 2010], Douban [Ma et al., 2011], and Movielens-100K [Harper and Konstan, 2016]. For Flixster, we show results for both user/item graphs (right number) and user graph only (left number). Baseline numbers are taken from [Monti et al., 2017, Berg et al., 2017].

weighed with the appropriate weights and minimized with respect to $\boldsymbol{C}$, $\boldsymbol{P}$ and $\boldsymbol{Q}$ by gradient descent,

$$\min_{\boldsymbol{C},\boldsymbol{P},\boldsymbol{Q}} E_{\mathrm{z}} + \mu_r E^r_{\mathrm{dir}} + \mu_c E^c_{\mathrm{dir}} + o_r E^r_{\mathrm{orth}} + o_c E^c_{\mathrm{orth}} + \rho_r E^r_{\mathrm{diag}} + \rho_c E^c_{\mathrm{diag}}. \tag{6}$$

4 Results and discussion

We demonstrate the effectiveness of our approach on the following datasets: Synthetic Netflix, Flixster, Douban and Movielens (ML-100K) as referenced in Table 1. The datasets include user ratings for items (such as movies) and additional features. For all the datasets we use the users and items graphs taken from Monti et al. [2017]. See Table 2 in the Appendix for a summary of the dataset statistics. The compared methods are referenced in Table 1. Two variants of the proposed spectral geometric matrix completion method (SGMC) are evaluated. For these methods we choose a maximal resolution $p_{\max}, q_{\max}$ (which can be larger than m, n) and a skip determining the spectral resolution (i.e., which $w_{p,q}$ in (5) are set to 1). SGMC uses only $w_{p_{\max},q_{\max}} = 1$, with the rest set to zero, while SGMC-Z uses $w_{1+kp_{\mathrm{skip}},1+kq_{\mathrm{skip}}} = 1,\ k \in \mathbb{N}$. These and other hyper parameters for our algorithms were tuned by hand, and are summarized in Table 3 in the Appendix. We shall provide a thorough investigation of their effect in an extended version of this article. The method labeled DMF minimizes the loss $\left\|\left(\boldsymbol{P}\boldsymbol{C}\boldsymbol{Q}^\top - \boldsymbol{M}\right) \odot \boldsymbol{S}\right\|_F^2$, i.e., it does not incorporate any geometric side information. This method coincides with the one suggested by Arora et al. [2019].

A few remarkable observations can be extracted from Table 1: First, on the Douban and ML-100K dataesets, the simple DMF shows competitive performance with all the other methods. This suggests that the geometry information is not very useful for these datasets. Second, the proposed algorithm outperforms the other methods, despite its simple and fully linear architecture. This suggests that the other methods do not exploit the geometry properly, and this fact is obscured by their cumbersome architecture. Third, while some of the experiments reported in Table 1 showed only slight margins in favor of SGMC/SGMC-Z compared to DMF, the results in the Synthetic Netflix column suggest that when the geometric model is accurate our methods demonstrate superior results. We further validated this on a synthetic model generated from the ML-100k graphs. These experiments will be reported in an extended version of this article.

5 Conclusion

In this work we have proposed a simple spectral technique for matrix completion, extending ideas borrowed from the field of non-rigid shape analysis. Our approach combines a full multiresolution spectral loss with (implicit) metric learning. Under a suitable change of basis, we obtain a fully linear network that gives rise to a useful interpretation via the framework of functional maps. We have demonstrated state-of-the-art results on a few recommendation systems datasets, surpassing results obtained by much more complicated architectures. In addition, we have demonstrated that some results which are usually attributed to a clever use of geometry, can be obtained without any geometry altogether. We refer the reader to an extended version of this work for detailed explanation, additional results, theoretical discussions, and ablation studies.

References

Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717, 2009.

Sanjeev Arora, Nadav Cohen, Wei Hu, and Yuping Luo. Implicit regularization in deep matrix factorization. *arXiv preprint arXiv:1905.13655*, 2019.

Vassilis Kalofolias, Xavier Bresson, Michael Bronstein, and Pierre Vandergheynst. Matrix completion on graphs. *arXiv preprint arXiv:1408.1717*, 2014.

Federico Monti, Michael Bronstein, and Xavier Bresson. Geometric matrix completion with recurrent multi-graph neural networks. In *Advances in Neural Information Processing Systems*, pages 3697–3707, 2017.

Felix Wu, Tianyi Zhang, Amauri Holanda de Souza Jr, Christopher Fifty, Tao Yu, and Kilian Q Weinberger. Simplifying graph convolutional networks. *arXiv preprint arXiv:1902.07153*, 2019.

Xiao Peng Li, Lei Huang, Hing Cheung So, and Bo Zhao. A survey on matrix completion: Perspective of signal processing. *arXiv preprint arXiv:1901.10885*, 2019.

Rianne van den Berg, Thomas N Kipf, and Max Welling. Graph convolutional matrix completion. *arXiv preprint arXiv:1706.02263*, 2017.

Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (TOG)*, 31(4):30, 2012.

Maks Ovsjanikov, Etienne Corman, Michael Bronstein, Emanuele Rodolà, Mirela Ben-Chen, Leonidas Guibas, Frederic Chazal, and Alex Bronstein. Computing and processing correspondences with functional maps. In *SIGGRAPH ASIA 2016 Courses*, page 9. ACM, 2016.

Or Litany, Emanuele Rodolà, Alexander M Bronstein, and Michael M Bronstein. Fully spectral partial shape matching. In *Computer Graphics Forum*, volume 36, pages 247–258. Wiley Online Library, 2017.

Matthias Vestner, Roee Litman, Emanuele Rodolà, Alex Bronstein, and Daniel Cremers. Product manifold filter: Non-rigid shape correspondence via kernel density estimation in the product space. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3327–3336, 2017a.

Matthias Vestner, Zorah Lähner, Amit Boyarski, Or Litany, Ron Slossberg, Tal Remez, Emanuele Rodola, Alex Bronstein, Michael Bronstein, Ron Kimmel, et al. Efficient deformable shape correspondence via kernel matching. In *2017 International Conference on 3D Vision (3DV)*, pages 517–526. IEEE, 2017b.

Simone Melzi, Jing Ren, Emanuele Rodola, Maks Ovsjanikov, and Peter Wonka. Zoomout: Spectral upsampling for efficient shape correspondence. *arXiv preprint arXiv:1904.07865*, 2019.

Akshay Gadde, Sunil K Narang, and Antonio Ortega. Bilateral filter: Graph spectral interpretation and extensions. In *2013 IEEE International Conference on Image Processing*, pages 1222–1226. IEEE, 2013.

Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in neural information processing systems*, pages 109–117, 2011.

Nikhil Rao, Hsiang-Fu Yu, Pradeep K Ravikumar, and Inderjit S Dhillon. Collaborative filtering with graph information: Consistency and scalable methods. In *Advances in neural information processing systems*, pages 2107–2115, 2015.

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.

Hao Ma, Dengyong Zhou, Chao Liu, Michael R Lyu, and Irwin King. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 287–296. ACM, 2011.

F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):19, 2016. URL `https://grouplens.org/datasets/movielens/`.

6 Appendix

Dataset	Users	Items	Features	Ratings	Density	Rating levels
Flixster	3,000	3,000	Users/Items	26,173	0.0029	0.5, 1, ..., 5
Douban	3,000	3,000	Users	136,891	0.0152	1, 2, ..., 5
MovieLens-100K	943	1,682	Users/Items	100,000	0.0630	1, 2, ..., 5

Table 2: Number of users, items and ratings for Flixster, Douban and Movielens-100K datasets used in our experiments and their respective rating density and rating levels.

Dataset	$\mathrm{p}_{\max}/q_{\max}$	$\mathrm{p}_{\mathrm{skip}}/q_{\mathrm{skip}}$	μ_r/μ_c	ρ_r/ρ_c	o_r/o_c	Trainable variables
Synthetic	80/80	–/–	0.01/0.01	0/–	0/–	$\boldsymbol{P}, \boldsymbol{C}$
Netflix	500/500	3/1	0.4/0.4	0.1/–	0/–	$\boldsymbol{P}, \boldsymbol{C}$
Flixster	100/100	–/–	0.0025/0.0025	0.001/–	0/–	$\boldsymbol{P}, \boldsymbol{C}$
(both graphs)	100/100	1/1	0.0025/0.0025	0.001/–	0/–	$\boldsymbol{P}, \boldsymbol{C}$
Douban	2500/2500	–/–	0.001/0	0.001/0	0.001/0	$\boldsymbol{P}, \boldsymbol{C}, \boldsymbol{Q}$
	1000/1000	50/1000	0.011/0	0.004/0	0.001/0	$\boldsymbol{P}, \boldsymbol{C}, \boldsymbol{Q}$
ML-100K	5000/5000	–/–	0.0002 /0.0002	0.0002 /0.0002	0/0	$\boldsymbol{P}, \boldsymbol{C}, \boldsymbol{Q}$
	3200/3200	30/35	0.03/0.03	0.2/0.2	0.09/0.09	$\boldsymbol{P}, \boldsymbol{C}, \boldsymbol{Q}$

Table 3: Table of settings for the algorithms reported in Table (1). For each dataset the top row is SGMC and the bottom row is SGMC-Z.