Recurrent Time-Varying Multi-Graph Convolutional Neural Network for Personalized Cervical Cancer Risk Prediction

Vinay Chakravarthi Gogineni*, Severin R E Langberg†, Valeriya Naumova*, Jan F Nygård†,
Mari Nygård‡, Markus Grasmair§, and Stefan Werner‖
*Dept. of Machine Intelligence, Simula Metropolitan Center for Digital Engineering, Norway
†Dept. of Registry Informatics, Cancer Registry of Norway, Norway
‡Dept. of Research, Cancer Registry of Norway, Norway
§Dept. of Mathematical Sciences, Norwegian University of Science and Technology, Norway
‖Dept. of Electronic Systems, Norwegian University of Science and Technology, Norway.

***Abstract*—Cervical cancer screening programs have reduced the incidence of cervical cancer, but suffer from over- and too infrequent screening as women's risk of developing cervical cancer differs. Personalized risk prediction models contribute toward efficient, personalized cancer screening. This paper presents a personalized time-dependent cervical cancer risk prediction scheme to aid experts in recommending screening intervals. From partially observed screening histories, the proposed approach learns time-varying row-graphs that model the time-varying relations among the screening records of patients and a column-graph that encodes smoothness of an individual screening history. Then, leveraging these geometric structures, we reconstruct the entire latent risk of each individual from scarce screening data. In order to accomplish this, a novel time-varying multi-graph convolution neural network is proposed. These estimated latent risk profiles are used to forecast the cancer risk of new patients. The proposed approach is tested both on synthetic and real-life screening data obtained from the Cancer Registry of Norway.**

I. Introduction

The human papillomavirus (HPV) causes cervical cancer, which develops cellular changes, from low-grade lesions to high-grade (pre-cancerous) lesions to invasive cancer [1]. For Norwegian women aged 25 to 49, cervical cancer ranks third among the most common types of cancer. According to estimates, approximately 1% of Norwegian women will develop cervical cancer by age 75 [2]. Cervical cancer screening tests such as Cytology, Histology, or HPV can predict a woman's risk of developing cervical cancer [3]. Nordic countries have implemented mass-screening programs to detect and prevent cervical cancer in females [4]. The Norwegian Cervical Cancer Screening Program (NCCSP) recommends regular screening every third year starting at age 25 and ending at age 69, which means 15 screenings if they are all normal [5]. The risk of being infected with HPV and developing cancer differs greatly between females and also over time. As a result, regular cervical cancer screening leads to over-screening (i.e., unnecessary screenings for patients unlikely to develop the disease) or infrequent screening (i.e., very few screenings for patients at high risk) problems [6]. A personalized screening interval can be inferred as a solution by predicting the future risk of cervical cancer development. To this end, leveraging the relations between screening histories of patients, a personalized risk prediction scheme has been proposed in [7]. As the risk of being infected with HPV varies over time, the relations among screening histories of females will also change over time. Thus, it makes sense to model the relations among the screening histories of females using time-varying graphs. In this work, we develop a time-dependent cancer risk prediction scheme, which capitalizes on the time-varying relationship between screening histories. Our contributions are as follows:

- During training, under the assumption of *temporal homogeneity* [8], time-varying row-graphs are learned from partially observed screen histories. A column-graph is also generated under the assumption that cervical cancer risk does not change dramatically within a year. Then, utilizing these geometric structures, continuous screening profiles are reconstructed from the partially observed screening histories using a novel recurrent time-varying multi-graph convolutional neural network.
- Inference is performed by using the matrix comprising estimated continuous risks of individuals, referred to as a dictionary. A given female's cancer development risk can then be predicted by computing and maximizing the probabilities of possible risk conditions.
- Both synthetic and real-world datasets obtained from the Norwegian Cancer Registry are used to examine the performance of the proposed approach.

II. Cervical Cancer Screening Data

Data collected through NCCSP includes the results of three clinical examinations, namely, cytology, histology, and HPV, as well as the date of these examinations. These screening test results are labeled into four risk states: *normal*, *low-risk*, *high-risk* and *cancer* that indicate how likely it is that one will develop cervical cancer [9]. The *normal* state represents an accepted baseline risk. The *low-risk* state indicates an

This work was supported by the Research Council of Norway.

early stage of carcinogenesis, hence frequent screening is recommended to detect the disease before it becomes invasive, but there is usually no need for treatment. An *high-risk* state denotes a high likelihood of future cancer progression and requires immediate treatment. Finally, a *cancer* state is a result of a screening program failure and is a potential state of a patient.

Fig. 1. Cervical cancer screening histories of patients arranged in a matrix.

The cervical cancer screening data is arranged in a matrix $\mathbf{X} \in \mathbb{N}^{S\times N}$, where $X_{s,n} \in \{1,2,3,4\}$ is an integer encoding the observed states in the screening exam. The rows of $\mathbf{X}$ represent the partially observed screening history of a patient, and the column represents the age at which the observations were made. According to NCCSP, effective screening interval for healthy patients is 3 years, and for low-risk patients is 3-6 months [10], so a 3-month interval is then considered for data discretization. As shown by Fig. 1, there are only a few entries in the screening data matrix due to the sparse cervical screenings. The screening histories of individuals are also irregular as the recommendations are not strictly followed. Additionally, the screening results are highly skewed, meaning that most of the screening results are normal. Here the objective is to develop a cervical cancer risk prediction model from this highly challenging screening data.

III. Personalized Cervical Cancer Risk Prediction

A. Learning the Graphs

For the graph-based methods to be used for reconstructing the continuous latent risk of females, we must supply both row- and column-graphs. Graphs that encode the relations between screening histories are not provided with the screening data, so they must be inferred from partially observed screening data. To infer network/graph topology from the data, numerous techniques have been proposed in the literature [11]–[13]. As the risk of being infected with HPV changes over time, the relationships between screening histories may change as well. Therefore, rather than a single row-graph, we propose to use time-varying row-graphs for encoding/modelling the time-varying relations between screening histories of patients.

In this work, we learn time-varying graphs under the assumption of *temporal homogeneity*, i.e., most of the edges and their weights of time-varying graph remain unchanged over a short period of time [8]. In other words, certain females in the population exhibit similar screening histories within a short period of time. For this purpose, we first partition the data matrix $\mathbf{X}$ into T non-overlapping windows, where every window $\mathbf{X}_t$ for $t = 1, 2, \cdots, T$, covers K number of time points (i.e., screening results) of each screening history. Then, time-varying row-graphs can be obtained by solving the following fused LASSO problem [8]:

$$\min_{\mathbf{W}_{r,t}t\in\mathcal{W}_m} \sum_{t=1}^{T} \frac{1}{2}\|\mathbf{W}_{r,t} \circ \mathbf{Z}_t\|_1 + f(\mathbf{W}_{r,t}) + \eta \sum_{t=2}^{T} \|\mathbf{W}_{r,t} - \mathbf{W}_{r,t-1}\|_1, \quad (1)$$

where $\mathbf{Z}_t$ is the pairwise distance matrix defined by

$$[\mathbf{Z}_t]_{i,j} = \sum_{k=1}^{K} \|\mathbf{x}^i_{k,t} - \mathbf{x}^j_{k,t}\|^2, \quad (2)$$

with $\mathbf{x}_{k,t}$ is the kth column of the data window $\mathbf{X}_t$. The weighted adjacency matrix of the row-graph at time t is denoted by $\mathbf{W}_{r,t}$ and $\mathcal{W}_m$ is the space that contains all valid non-negative, symmetric weighted adjacency matrices, i.e., $\mathcal{W}_m = \{\mathbf{W}_{r,t} \in \mathbb{R}_+^{S\times S} : \mathbf{W}_{r,t} = \mathbf{W}_{r,t}^{\mathrm{T}}, \mathrm{diag}(\mathbf{W}_{r,t}) = 0\}$. The regularization function $f(\mathbf{W}_{r,t})$ in (1) prevents $\mathbf{W}_{r,t}$ being a zero matrix. Researchers used various functions for $f(\mathbf{W}_{r,t})$ in the literature. However, we follow [12] to obtain the sparse solution which is important in the case of large scale applications, and use $f(\mathbf{W}_{r,t}) = -\alpha \mathbf{1}^{\mathrm{T}} \log(\mathbf{W}_{r,t}\mathbf{1}) + \beta\|\mathbf{W}_{r,t}\|_{\mathrm{F}}^2$ with $\alpha \geq 0$ and $\beta \geq 0$. The logarithmic barrier forces the node degrees to be positive and the parameter β helps to control the sparsity of the graph, i.e., as β decreases, the solution of (1) becomes more sparse. Finally the third term, which is the difference between neighboring time windows, promotes the temporal homogeneity. We used primal-dual techniques [14] for solving the optimization problem stated in (1). Finally, we construct the column-graph under the assumption that the risk of cancer development does not change rapidly within a year.

B. Reconstructing the Latent Risk of Cervical Cancer

A patient's observed state $X_{s,n}$, is considered to be a noisy measurement of underlying latent risk, $Y_{s,n}$, that slowly evolves over time. Specifically, we assume that observed states are derived from a discrete Gaussian distribution with mean $Y_{s,n}$ and variance $\frac{1}{2\theta}$, with $\theta > 0$. By employing the principles of geometric matrix completion [15], [16], we aim to estimate the continuous latent risk $\mathbf{Y}$ from the partially observed screen histories $\mathbf{X}$.

Let an undirected weighted row-graph $\mathcal{G}_r$, with associated graph Laplacian $\mathbf{L}_r = \mathbf{\Phi}\mathbf{\Lambda}_r\mathbf{\Phi}^{\mathrm{T}}$, encode the relationships between screening histories of patients. Similarly, an undirected weighted column-graph $\mathcal{G}_c$, with associated graph Laplacian

$\mathbf{L}_c = \mathbf{\Psi}\mathbf{\Lambda}_c\mathbf{\Psi}^{\mathrm{T}}$, encodes the temporal smoothness of an individual screening history. Then, leveraging on these geometric structures, the geometric matrix completion approaches [15], [16] reconstruct $\mathbf{Y}$ from $\mathbf{X}$ by solving:

$$\min_{\mathbf{Y}} \|\mathbf{P}_\Omega \circ (\mathbf{Y} - \mathbf{X})\|_{\mathrm{F}}^2 + \frac{\gamma_r}{2}\|\mathbf{Y}\|_{\mathbf{L}_r}^2 + \frac{\gamma_c}{2}\|\mathbf{Y}^{\mathrm{T}}\|_{\mathbf{L}_c}^2, \quad (3)$$

where the symbol $\circ$ is the Hadamard product operator and $\mathbf{P}_\Omega$ is an indicator matrix of observed entries set of $\mathbf{X}$. The regularization terms $\frac{\gamma_r}{2}\|\mathbf{Y}\|_{\mathbf{L}_r}^2$ and $\frac{\gamma_c}{2}\|\mathbf{Y}^{\mathrm{T}}\|_{\mathbf{L}_c}^2$ quantify the smoothness of all screening profiles over the row-graph and column-graph, respectively. The regularization coefficients $\gamma_r, \gamma_c > 0$. The recurrent multi-graph convolutional neural network (RMGCNN) framework [16] efficiently reconstructs the underlying continuous latent risk $\mathbf{Y}$ by solving the above optimization problem.

Since the risk of being infected with HPV varies over time, the relations among the screening histories will also vary over time. Single row-graph fails in modeling these time-varying relationships between patients screening histories. So, we aim to use time-varying row-graphs to encode these time-varying relations, and used these time-varying geometric structures for reconstructing the continuous latent risk $\mathbf{Y}$. Using the time-varying row-graphs, the geometric matrix completion problem becomes:

$$\min_{\mathbf{Y}} \|\mathbf{P}_\Omega \circ (\mathbf{Y} - \mathbf{X})\|_{\mathrm{F}}^2 + \frac{\gamma_r}{2}\sum_{t=1}^{T}\|\mathbf{Y}_t\|_{\mathbf{L}_{r,t}}^2 + \frac{\gamma_c}{2}\|\mathbf{Y}^{\mathrm{T}}\|_{\mathbf{L}_c}^2, \quad (4)$$

To solve the above time-varying geometric matrix completion problem, we propose a novel recurrent time-varying multi-graph convolution neural network (RtvMGCNN), whose architecture is illustrated in Fig. 2.

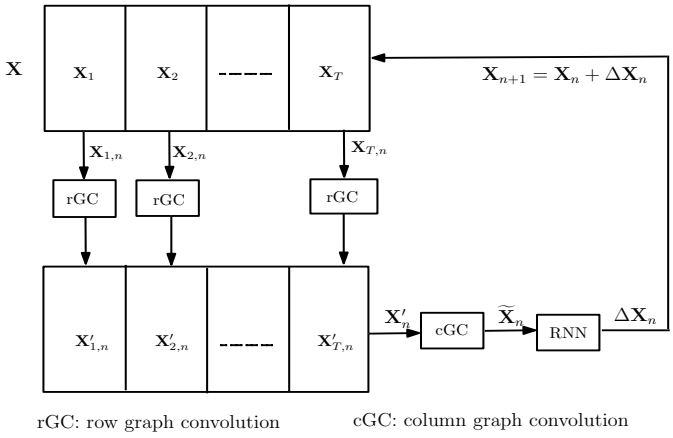

Fig. 2. RtvMGCNN architecture.

The RtvMGCNN solves the time-varying geometric matrix completion problem as follows: First the spatial features are extracted by performing the multigraph convolution using time-varying row-graphs and column graph as follows:

$$\widetilde{\mathbf{X}} = \sigma\Big(\sum_{j,j'=0}^{p} \theta_{j,j'}\ \mathsf{Append}\left\{\overline{\mathbf{L}}_{r,t}\mathbf{X}_t\right\}_{t=1}^{T}\ \overline{\mathbf{L}}_c\Big), \quad (5)$$

where $\theta_{j,j'}$ are the filtering coefficients in RtvMGCNN layer, and the function Append represents the appending operation. The matrices $\overline{\mathbf{L}}_r$ and $\overline{\mathbf{L}}_c$ be the respective Chebyshev polynomial of scaled Laplacians of the row- and column-graphs with eigenvalues are being in the interval $[-1, 1]$. In the next step, these extracted spatial features from tvMGCNN layer will be feeding to the recurrent neural network (RNN) that progressively reconstructs the complete screening profiles matrix by implementing a diffusion process. The RtvMGCNN uses an LSTM architecture to learn complex non-linear diffusion processes [16]. The tvMGCNN, together with LSTM predict accurately small changes of $\mathbf{X}$ that can propagate through the full temporal steps.

C. Predicting the Risk of Cervical Cancer Development

Given the screening record of a new patient $\mathbf{z}$, i.e., the screening results from $n_1, \cdots, n_k$, this section presents a method for predicting a patient's future state $z_{\hat{n}}$ for $\hat{n} > n_k$. By substituting empirical distribution of reconstructed latent risk $\hat{Y}_{s,\hat{n}}$ for the true distribution, we can calculate conditional probabilities for the future state $z_{\hat{n}}$ as follows:

$$\begin{aligned} p(z_{\hat{n}} = c \mid \mathbf{z}) \propto & \sum_{s=1}^{S} C_{\hat{Y}_{s,\hat{n}}} \exp(-\theta(c - \hat{Y}_{s,\hat{n}})^2) \\ & \times \prod_{j=1}^{k} C_{\hat{Y}_{s,n_j}} \exp(-\theta(z_{n_j} - \hat{Y}_{s,n_j})^2), \end{aligned} \quad (6)$$

where $C_{\hat{Y}_{s,n}}$ is a risk-dependent normalization constant and c denotes the state of screening result. Using (6), the conditional probabilities have to be calculated for $\forall c \in \{1, 2, 3, 4\}$. Then, the state with the highest conditional probability will be the predicted risk.

IV. Experimental Results

Numerical experiments were conducted on synthetic and real-life screening data to demonstrate the ability of the proposed RtvMGCNN in predicting the future risk of cervical cancer development. We generated synthetic data to resemble the high sparsity, randomness, and imbalance of the screening data. A latent risk matrix $\mathbf{Y} = \mathbf{U}\mathbf{V}^{\mathrm{T}}$ is synthesized from a rank-five basis of the form $V_{t,k} = \exp(-10^3(t - \mu_k)^2)$ with $\mu_k \in \{70, 95, 120, 145, 170\}$ and the patient-specific coefficients drawn from an exponential distribution. We obtain a partially observed matrix $\mathbf{X}$ so to resemble the scarcity, irregularity, and skewness of the NCCSP data. We repeat the procedure of synthesizing six datasets of each having 10000 samples with similar density (i.e., the fraction of observed entries in $\mathbf{X}$) for five different random seeds. Each dataset was partitioned into 80% training and 20% test samples. The trained data matrix was partitioned into 4 non-overlapping windows and their corresponding row-graphs were learned. The continuous latent risk matrix reconstructed from training data is used to estimate the conditional probabilities, which have been used to predict the future risk at specific time points in independent test data. The hyperparameters were optimized through cross-validation. The k-category Matthews correlation

coefficient (MCC) [17] that summarizes the confusion matrix by a number $\text{MCC}_k \in [-1, 1]$ was considered a performance metric. For comparative assessment, the task of predicting the future risk of cancer development has also been carried out by forward fill (FF) (in which the last screening result is repeated to fill the missing screening result) and RMGCNN [7] approaches. The performance of these models is compared to an Oracle, which returns the most likely screening result given the true latent risk $Y_{s,n}$. The MCC_k scores of various models vs. dataset density are illustrated in Fig. 3.

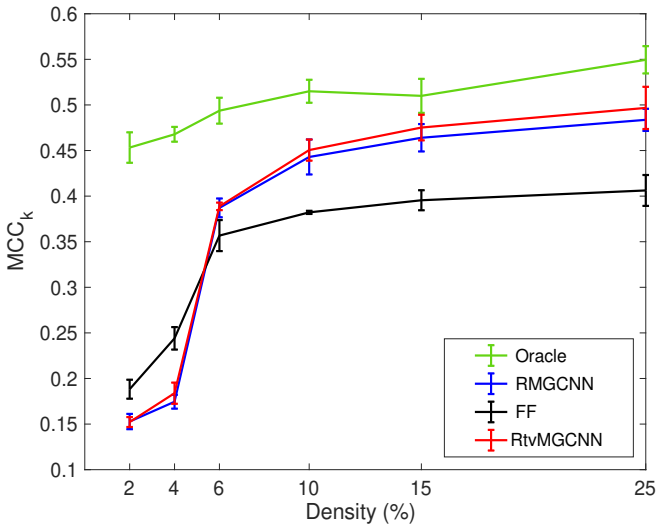

Fig. 3. Performance of proposed RtvMGCNN on synthetic data given as the K-category Matthews correlation coefficient (MCC_k) against dataset density. Also illustrated the performance of Oracle, FF and RMGCNN.

From Fig. 3, we see that the performance of all approaches is proportional to the synthetic dataset density. Above 6% dataset density, as the time-varying graphs efficiently modeled the relations among screening histories, the proposed RtvMGCNN based approach exhibited better performance on synthetic datasets compared to the FF, and RMGCNN. The time-varying graphs, on the other hand, failed to model the relationships at low dataset densities, hence, the proposed approach performs similar to RMGCNN.

A. Results on Real-life Screening Data

From the NCCSP population-level data, we randomly selected the data of 10000 female patients in which every female patient has at least one screening exam result. In this dataset, screening exams account for an average of 8 and the maximum is 37 per patient; this corresponds to 2.3% observed entries. With a temporal resolution of three months, each history was aligned over a time grid spanning from youngest to oldest screened patient. We used the last observation when multiple screenings occurred within a three month period in order to reflect the clinical data available to clinicians. In Table. I, we present the MCC_k scores for predicting future cancer development risk from NCCSP data one to three years in the future.

TABLE I
PERFORMANCE OF VARIOUS APPROACHES GIVEN AS THE K-CATEGORY MATTHEWS CORRELATION COEFFICIENT (MCC_k) ON THE NCCSP DATA

Forecast (years)	FF	MF	RMGCNN	RtvMGCNN
1	0.1505	0.1250	0.1649	0.1821
2	0.0804	0.0728	0.1407	0.1563
3	0.0834	0.0429	0.1215	0.1488

From Table. I, one can see that the RtvMGCNN performs slightly better than the FF, matrix factorization and RMGCNN approaches. However, all these methods exhibit poor performance on real-life screening data. This is due to low density of observed entries in the real-life screening dataset. When the real-life screening data matrix was partitioned to 4 non-overlapping windows for learning the time-varying row graphs, few data windows contained only 1% observed entries. Due to this, the time-varying row-graphs were unable to encode the relations among the screening histories in those windows.

V. CONCLUSIONS

In this paper, we considered the problem of predicting the future risk of cervical cancer development in an individual. For this, leveraging the time-varying relations among screening histories of patients continuous latent risk have been constructed. To this end a novel recurrent time-varying multi graph convolutional neural network has been proposed. The reconstructed cancer screening data matrix was then used to forecast the cancer risk of a new patient. The proposed approach has been tested on the synthetic and real-life datasets to demonstrate its potential. The numerical results revealed that the proposed approach can predict individuals' short-term risk of being diagnosed with cervical cancer (12-36 months), targeting those who would benefit from more frequent screenings in order to reduce under-treatment.

REFERENCES

[1] P. A. Cohen, A. Jhingran, A. Oaknin, and L. Denny, "Cervical cancer," *Lancet*, vol. 393, no. 10167, pp. 169–182, 2019.

[2] Cancer Registry of Norway, "Cancer in Norway 2018 - Cancer incidence, mortality, survival and prevalence in Norway," *Cancer in Norway*, 2019.

[3] WHO. Cervical cancer. [Online]. Available: https://www.who.int/health- topics/cervical-cancer

[4] S. Vaccarella, S. Franceschi, G. Engholm, S. Lönnberg, S. Khan, and F. Bray, "50 years of screening in the Nordic countries: Quantifying the effects on cervical cancer incidence," *Br. J. Cancer*, vol. 111, no. 5, pp. 965-969, Aug. 2014.

[5] K. Pedersen, S. Fogelberg, L. H. Thamsborg, M. Clements, M. Nygard, I. S. Kristiansen, E. Lynge, P. Sparen, J. J. Kim, and E. A. Burger, "An overview of cervical cancer epidemiology and prevention in Scandinavia," *Acta obstetricia et gynecologica Scandinavica*, vol. 97, no. 7, pp. 795–807, 2018.

[6] K. Pedersen, E. A. Burger, S. Campbell, M. Nygard, E. Aas, and S. Lonnberg, "Advancing the evaluation of cervical cancer screening: Development and application of a longitudinal adherence metric," *The European J. Public Health*, vol. 27, no. 6, pp. 1089–1094, 2017.

[7] V. C. Gogineni, G. S. R. E. Langberg, V. Naumova, J. Nygård, M. Nygård, M. Grasmair and S.Werner "Data-driven personalized cervical cancer risk prediction: A graph-perspective," in *Proc. IEEE Int. Workshop Statistical Signal Process.*, 2021 (Submitted).

[8] K. Yamada, Y. Tanaka and A. Ortega, "Time-varying graph learning based on sparseness of temporal variation," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, 2019, pp. 5411-5415.

[9] B. C. Soper, M. Nygard, G. Abdulla, R. Meng, and J. F. Nygard, "A hidden Markov model for population-level cervical cancer screening data," *Statistics in Medicine*, vol. 39, no.25, pp. 3569– 3590, 2020.

[10] J. F. Nygard, S. O. Thoresen, and G. B. Skare, "The cervical cancer screening program in Norway, 1992-2000: Changes in pap-smear coverage and cervical cancer incidence," *Int. J. Cancer*, pp. 110–110, 2002.

[11] G. Mateos, S. Segarra, A. G. Marques and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Processing Magaz.*, vol. 36, no. 3, pp. 16-43, May 2019.

[12] V. Kalofolias, "How to learn a graph from smooth signals," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2016, pp. 920–929.

[13] X. Dong, D. Thanou, P. Frossard and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160-6173, Dec. 2016.

[14] N. Komodakis and J. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving largescale optimization problems," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 31–54, 2015.

[15] V. Kalofolias, X. Bresson, M. M. Bronstein, and P. Vandergheynst, "Matrix completion on graphs," 2014. [Online]. Available: https://arxiv.org/abs/1408.1717

[16] F. Monti, M. Bronstein, and X. Bresson, "Geometric matrix completion with recurrent multi-graph neural networks," in *Proc. Advances in Neural Info. Process. Systems*, 2017, pp. 3697–3707.

[17] J. Gorodkin, "Comparing two k-category assignments by a k-category correlation coefficient," *Computational Biology and Chemistry*, vol. 28, no. 5-6, pp. 367–374, 2004.