

# PREDICTION OF LANGUAGE IMPAIRMENTS IN CHILDREN USING DEEP RELATIONAL REASONING WITH DWI DATA

Soumyanil Banerjee<sup>1</sup> Ming Dong<sup>1</sup> Min-Hee Lee<sup>2,4</sup> Nolan O'Hara<sup>4</sup> Eishi Asano<sup>2,3</sup> Jeong-Won Jeong<sup>2-4</sup>

Department of Computer Science<sup>1</sup>, Pediatrics<sup>2</sup>, Neurology<sup>3</sup>, Wayne State University, Detroit, USA  
Translational Imaging Laboratory<sup>4</sup>, Childrens Hospital of Michigan, Detroit, USA

## ABSTRACT

This paper proposes a new deep learning model using relational reasoning with diffusion-weighted imaging (DWI) data. We investigate how effectively and comprehensively DWI tractography-based connectome predicts the impairment of expressive and receptive language ability in individual children with focal epilepsy (FE). The proposed model constitutes a combination of a dilated convolutional neural network (CNN) and a relation network (RN), with the latter being applied to the dependencies of axonal connections across cortical regions in the whole brain. The presented results from 51 FE children demonstrate that the proposed model outperforms other existing state-of-the-art algorithms to predict language abilities without depending on connectome densities, with average improvement of up to 96.2% and 83.8% in expressive and receptive language prediction, respectively.

## 1. INTRODUCTION

Focal epilepsy (FE) disrupts brain functions supporting language development and increases the risk of intellectual disabilities [1, 2]. Given the importance of age-appropriate language development for educational and social wellbeing, it is clinically imperative that any impairments in language function are readily, objectively, and reliably identified. Indeed, there is an urgent need for whole-brain neuroimaging tools which can non-invasively identify altered brain networks underlying language impairment in FE, because neural dynamics supporting language function involve extensive brain networks not limited to a single region [3].

The overall purpose of this study is to determine how effectively and comprehensively diffusion-weighted imaging (DWI) tractography-based connectome can predict the impairment of expressive and receptive language ability in individual FE children. The first analytic step in this study was to model the whole brain as a large distributed network called a “connectome matrix,” represented by a collection of nodes (*i.e.*, cortical and subcortical regions) and edges (*i.e.*, pair-wise connections between nodes). By tracking a diffusion signal between two given nodes, we quantified pair-wise connectivity for each edge and then investigated the relationship of the quantified connectivity with expressive and receptive language scores as rated by the Clinical Evaluation of Lan-

guage Fundamentals (CELF) assessment [4].

As one of the most powerful deep learning models, convolutional neural networks (CNN) have been widely used in biomedical imaging tasks [5]. Recent studies have applied CNN to predict cognitive and neuromotor outcomes from connectivity networks in infant brains [6]. However, the convolutional masks are applied through a sliding window in CNN, and thus the extracted features are inevitably local without any long-range connectivity information. More importantly, the spatial arrangement of nodes is arbitrarily determined by a fixed cortical atlas (*e.g.*, Automated Anatomical Labelling (AAL) atlas [7]). This configuration leaves local edges ill-posed to represent specific functional networks, and creates a need for new deep learning networks that can feasibly consider non-local relations across more remote edges. Our working hypothesis is that straightforward analysis (*e.g.*, regression) based on local CNN features would be naturally limited - unable to identify and leverage the complex, non-local connectivity patterns needed to accurately predict language impairment.

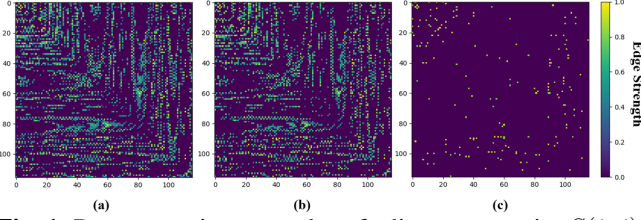
As an alternative approach, the present study proposes to predict language impairment using a dilated CNN augmented with a relation network (dilated CNN+RN). A relation network (RN) [8] is a special graph network [9] whose computations focus explicitly on relational reasoning. The main contributions of our work is summarized as follows:

1. To our knowledge, this is the first work that predicts language impairment using deep relational reasoning. Our predictive model, dilated CNN+RN, can identify and leverage the complex, non-local connectivity patterns in the connectome matrix obtained from clinically acquired DWI data.
2. Dilated CNN+RN significantly outperforms the current state-of-the-arts in predicting expressive and receptive language scores. Analysis of predicted and actual scores shows that dilated CNN+RN provides an effective tool to improve our understanding of the neuroanatomical substrates of language impairment in FE children.

## 2. METHODOLOGY

### 2.1. Study subjects

The present study included 51 children clinically diagnosed with FE (age:  $11.8 \pm 3.1$  years, 26 boys). Language ability



**Fig. 1.** Representative examples of adjacency matrix,  $S(i, j)$ , in DWI connectome. (a) high density ( $q=0.1$ ). (b) medium density ( $q=0.4$ ). (c) low density ( $q=0.8$ ).

was assessed using the CELF-Preschool (CELF-P) for children aged 2-5 years, and CELF-4th version for children aged 6 years and above. Composite measures derived from CELF yielded expressive and receptive language scores that were standardized to a mean of 100 and standard deviation of 15, with a sample mean  $\pm$  standard deviation of  $76 \pm 24$  and  $81 \pm 24$  in expressive and receptive language ability, respectively. These scores were used as the ground-truth to generate our RN-based predictive models using DWI connectome. The study was approved by University's Institutional Review Board (IRB).

## 2.2. Construction and augmentation of DWI connectome

The Automated anatomical labeling (AAL) parcellation scheme [7] was spatially normalized to a T1-weighted image using Advanced Normalization Tools [10] and then applied to whole brain tractography in order to construct a DWI-based brain connectome, where a set of 116 nodes represents regions of interest in the whole brain, and  $S$  is an adjacency matrix of edges representing the strength of pair-wise connection between each pair of brain nodes. The number of white-matter streamlines connecting the nodes is normalized by both average streamline length and volume of the nodes in order to stabilize inter-subject variability (due to intracranial volume and age differences). To reduce false positives, we identified true positive connections of  $S$  as pair-wise connections for which the strength values are greater than the threshold of quantiles satisfying  $q$  cumulative probability of element value in  $S$  (i.e.,  $q = 0.1/0.4/0.8$  for high, medium and low density, respectively, as shown in Fig. 1).

To prevent overfitting of RN layers in a relatively small cohort ( $n=51$ ), we applied synthetic minority over-sampling technique (SMOTE) [11] which enlarges the training and testing dataset up to 510 augmentations per patient. Each  $[S_m, t_m]$  of the  $m^{th}$  study subject was stacked in a  $51 \times 6729$  matrix, where the  $i^{th}$  row is a  $1 \times 6729$  vector (6728 elements of  $S_m$  and a scalar of  $t_m$  i.e., CELF score). Then each row of this matrix was augmented 510 times by randomly interpolating its 6 nearest neighbors, resulting in an augmented matrix of size  $26061 \times 6729$ . We randomly assigned 34 patient data and their augmented counterparts as a training cohort which comprised of 17,374 instances (66.7%). The remaining 17 patient data and their augmented counterparts were assigned into a test cohort which comprised of 8,687

instances (33.3%).

## 2.3. Dilated CNN+RN for prediction of language score

The proposed dilated CNN+RN performs relational reasoning with the connectome matrix. Since our input image (the connectome matrix) is of size  $116 \times 116$ , if we consider the pixels of the image as nodes of a graph network, the number of nodes in the graph network would be  $V = (116)^2$  and the number of edges would be  $E = \frac{V(V-1)}{2}$ , making the graph network practically non-trainable. Hence, we need a function that can map the pixels of the image to the nodes of the graph network. This function is achieved by the convolutional operation in CNN. Notice that the convolution operation of a kernel only covers a local spatial region of the image. Here, we need an inference model which can leverage non-local relations between the extracted features.

### 2.3.1. Relational Network

Fig. 2 presents the architecture of the dilated CNN+RN model where CNN is used to parse the pixel inputs into a set of features. The output of CNN is a set of feature maps and an object is a slice of the feature map through the third dimension. RN [8] focuses explicitly on relational reasoning to capture the dependencies of axonal connections: *edges* (i.e., objects) across cortical regions of the whole brain. Mathematically, RN is a composite function given by:

$$t_m = RN(O) = f_\phi \left( \frac{1}{N} \sum_{i,j} g_\theta(o_i, o_j) \right) \quad (1)$$

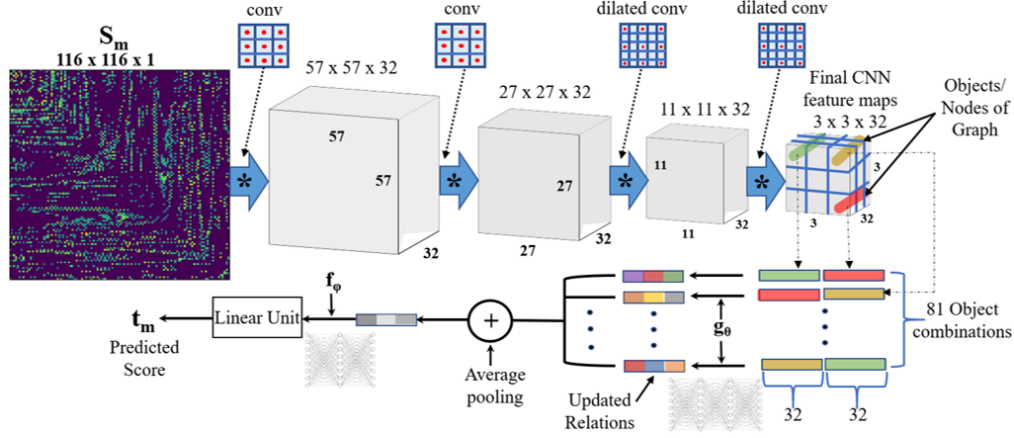
where  $o_i$  and  $o_j$  are a possible object pair obtained from the feature maps; the function  $g$  is a fully connected (FC) network which operates on these object pairs and computes relations between them; and the function  $f$  is another FC network which operates on the averaged relations and predicts the final score.  $N$  is the total number of relations obtained from the final feature maps.

### 2.3.2. Dilated convolution

Since the connectome matrix was sparse, and we wanted to consider all the possible relations between extracted objects, we used dilated convolution kernels in CNN [12], which have a larger receptive field than usual kernels, resulting in a smaller feature map and thereby obtaining more important relation vectors to be used by the RN. The dilated convolution operation is given as,

$$(x *_l h)(p) = \sum_{s+lt=p} x(s)h(t) \quad (2)$$

where  $l$  is the dilation factor,  $x$  is the image and  $h$  is the convolution kernel. For the convolution operation, if the input image shape is  $X$ , kernel size is  $f$ , padding is  $p$ , stride is  $s$ ,



**Fig. 2.** Network architecture of the proposed dilated CNN+RN which takes a given input  $S_m$  to predict an output score  $t_m$ .

dilation factor is  $l$ , then the output shape  $D$  is given by:

$$D = \left\lceil \frac{X - f - (f - 1) * (l - 1) + 2 * p}{s} \right\rceil + 1 \quad (3)$$

### 2.3.3. Gradient based activation maps

We adopted a generalized gradient-based class activation mapping (Grad-CAM - [13]) in dilated CNN+RN to understand which connections in the connectome matrix are important for the final prediction. In Grad-CAM, the weights of the gradients  $\alpha_k^c$  are first calculated by taking the partial derivative of output score w.r.t the set of feature maps as shown below,

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y_c}{\partial A_{ij}^k} \quad (4)$$

where  $y_c$  is the score for class  $c$ ,  $A^k$  is the set of final feature maps after the last layer of convolution, and  $Z$  is a normalizing factor. The final activation map  $L_{Grad-CAM}^c$  is given as follows:

$$L_{Grad-CAM}^c = RELU\left(\sum_k \alpha_k^c A^k\right) \quad (5)$$

where the RELU operation removes the negative gradients.

In our architecture, we computed the activation maps after the fourth convolution layer. We used the regression score after the final linear layer to obtain the activation maps for the CELF expressive and receptive scores.

## 2.4. Implementation

Due to the symmetric property of the connectome matrix, we used only the upper triangular part of the matrix as the input to our model. The output after the final dilated CNN layer were 32 feature maps each of size  $3 \times 3$ . An object was sliced from the third dimension of the feature maps as shown in Fig. 2. The final layer of the network following  $f$  was a linear layer with one unit which was required for regression.

The model was trained for 1000 epochs with adam optimizer having learning rate of 0.0001 and with a batch size of 128. Mean squared error was used as the loss function.

## 3. RESULTS

### 3.1. Experiment setup

Computational experiments were carried out to compare the performance of the dilated CNN+RN model with current state-of-the-art models used for prediction of language impairment. In the same training and testing splits, we compared the prediction performances of the following models:

**Baseline:** Multi-Layer Regressor (MLR) - comprised of a four layer fully connected network with 512 units in the first 2 layers followed by 256 units in the third layer and a single linear unit in the final layer.

**CNN models:** CNN+MLR - comprised of a four layer CNN where each layer of CNN had 32 kernels of size  $3 \times 3$ , followed the same MLR as the baseline.

**State-of-the-art model:** BrainNetCNN [6] (Edge-to-Node Network, E2N) - E2N layer had 32 cross-shaped kernels with size equal to the input connectome matrix, i.e.,  $(116 \times 1) \times (1 \times 116)$ , which were strided along the diagonal of the connectome matrix. This was followed by a fully connected network with 30 units.

**Our proposed model:** Dilated CNN+RN - comprised of a dilated CNN with four layers, similar to the CNN model described above, but with the last two layers having a dilated kernel with dilation factor of 2. The RN comprised of function  $g$  which had four fully connected layers, each layer having 512 units. The output of  $g$  was averaged and then the function  $f$  was applied on this combination. Function  $f$  comprised of three fully connected layers with 512 units. The final layer was a linear layer.

Three metrics were used to evaluate the performance of each model, including the mean absolute error (MAE) between the observed and actual receptive and expressive scores [6], standard deviation of absolute error (SDAE) [6],

**Table 1.** Performance comparison of the dilated CNN+RN with other models. For each model, MAE, SDAE and PE were evaluated at 3 different connectome densities: density 1 (high), density 2 (medium), and density 3 (low). R and E indicate receptive score and expressive score, respectively.

Network	Score Type	Density 1 (MAE, SDAE, PE)	Density 2 (MAE, SDAE, PE)	Density 3 (MAE, SDAE, PE)
MLR	R	26.796, 12.696, 14.26%	25.068, 12.168, 19.12%	14.232, 8.7720, 57.85%
	E	26.352, 12.852, 24.17%	27.120, 13.284, 24.05%	16.092, 10.692, 48.90%
CNN + MLR	R	3.7200, 2.7360, 100.0%	3.1800, 2.5320, 100.0%	8.0400, 5.6760, 94.15%
	E	1.8720, 1.6320, 100.0%	1.9680, 1.5000, 100.0%	11.640, 6.5520, 70.62%
E2N	R	1.2000, <b>1.0800</b> , 100.0%	1.4520, <b>1.3680</b> , 100.0%	14.196, 8.8200, 59.13%
	E	1.6920, 1.3800, 100.0%	2.2800, 2.3040, 100.0%	16.776, 10.008, 46.73%
Dilated CNN + RN	R	<b>0.9120</b> , 1.5000, <b>100.0%</b>	<b>0.7660</b> , 1.7400, <b>100.0%</b>	<b>1.0440</b> , <b>2.7360</b> , <b>100.0%</b>
	E	<b>0.2120</b> , <b>0.2520</b> , <b>100.0%</b>	<b>0.2060</b> , <b>0.2440</b> , <b>100.0%</b>	<b>0.3720</b> , <b>0.5160</b> , <b>100.0%</b>

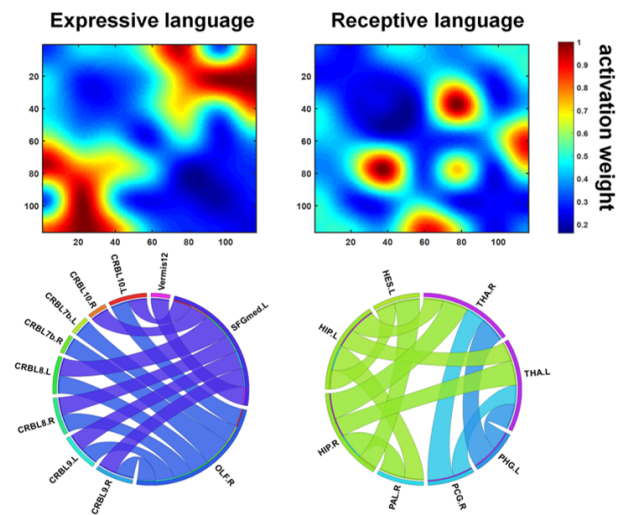
and probability of MAE less than 15 (PE) (*i.e.*, the frequency of the predicted score close to the actual score by a margin less than 15, one standard deviation of the normative cohort).

### 3.2. Prediction of language scores

Table 1 presents prediction performances of dilated CNN+RN and other models. The dilated CNN+RN yielded the best performance compared to other models. Both MAE and SDAE of the dilated CNN+RN are significantly lower than those of other methods when predicting expressive and receptive scores, yielding a respective 87-98% and 24-92% improvement in MAE over E2N. On average, MAE for expressive score prediction was reduced up to 6.65, with the greatest reduction of 16.4 (for lowest density) and the least reduction of 1.48 (for highest density). On average, MAE for receptive score prediction was reduced by 4.71, with the greatest reduction of 13.15 (for lowest density) and the least reduction of 0.29 (for highest density). These superior performances yielded 100% of PE to predict both expressive and receptive scores, suggesting that the dilated CNN+RN can provide a reliable means to predict language impairment, and ultimately may help supplement or replace portions of laborious neuropsychological assessments in clinical cases with serious behavioral problems.

### 3.3. Activation maps of language scores

Fig. 3 shows the activation maps (partial derivatives) associated with expressive and receptive language scores (Eq. 5), averaged over the top 20 subjects (with minimal MAE). 2-D circular connectograms also present multiple hub nodes having the highest partial derivatives (*i.e.*, Z-score of activation weight  $\geq 2.5$  and 3.5 for expressive and receptive language) and their connections to other nodes (*i.e.*, each line segment indicates a Z-score of activation weight). For prediction of expressive language score, left superior frontal medial cortex (SFGmed.L) and right olfactory cortex (OLF.R) appeared to be prominent hubs of important connections, as previously reported [14]. That is, connection edges across these hubs were found to be the most predictive of high expressive score. In contrast, bilateral hippocampus (HIP.L and HIP.R) and



**Fig. 3.** Activation maps showing AAL regions learned by the dilated CNN+RN as the most predictive of CELF scores. Each 2D circular connectome presents Circos ideogram available at <http://mkweb.bcgsc.ca/tableviewer/>.

left Heschl gyrus (HES.L) appear to be prominent hubs of important connections for receptive language score prediction [15]. Connection edges from hippocampus to thalamus (HIP.L-THA.L and HIP.L-THA.R) were the most predictive of receptive language score. Specifically, weaker connections across these regions were predictive of worse receptive language scores, suggesting severe language impairment. There were no noticeable overlaps between the edges that are most predictive of expressive and receptive scores.

## 4. CONCLUSION

The present study investigated the clinical utility of deep relational reasoning to predict language abilities from DWI connectome data of children with focal epilepsy. This approach may lead to the refinement of imaging and language phenotype relationships, and ultimately could help develop individualized therapeutic interventions for language impairment in young children, who are difficult to diagnose by current neuropsychological evaluations.

## 5. REFERENCES

- [1] Bruce P Hermann, Brian Bell, Michael Seidenberg, and Austin Woodard, "Learning disabilities and language function in epilepsy," *Epilepsia*, vol. 42, pp. 21–23, 2001.
- [2] Gillian M Parkinson, "High incidence of language disorder in children with focal epilepsies," *Developmental medicine and child neurology*, vol. 44, no. 8, pp. 533–537, 2002.
- [3] Yasuo Nakai, Ayaka Sugiura, Erik C Brown, Masaki Sonoda, Jeong-Won Jeong, Robert Rothermel, Aimee F Luat, Sandeep Sood, and Eishi Asano, "Four-dimensional functional cortical maps of visual and auditory language: Intracranial recording," *Epilepsia*, vol. 60, no. 2, pp. 255–267, 2019.
- [4] Teresa Paslawski, "The clinical evaluation of language fundamentals, (celf-4) a review," *Canadian Journal of School Psychology*, vol. 20, no. 1-2, pp. 129–134, 2005.
- [5] Ge Wang, Jong Chu Ye, Klaus Mueller, and Jeffrey A Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1289–1296, 2018.
- [6] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, "BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *Neuroimage*, vol. 146, pp. 1038–1049, 02 2017.
- [7] "Automated anatomical labeling single-subject atlas," *Montreal Neurological Institute (MNI)*.
- [8] Adam Santoro, David Raposo, David G Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap, "A simple neural network module for relational reasoning," in *Advances in neural information processing systems*, 2017, pp. 4967–4976.
- [9] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al., "Relational inductive biases, deep learning, and graph networks," *arXiv preprint arXiv:1806.01261*, 2018.
- [10] Brian B Avants, Nicholas J Tustison, Gang Song, Philip A Cook, Arno Klein, and James C Gee, "A reproducible evaluation of ants similarity metric performance in brain image registration," *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011.
- [11] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [12] Fisher Yu and Vladlen Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
- [13] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 618–626.
- [14] Satoshi Ookawa, Rei Enatsu, Aya Kanno, Satoko Ochi, Yukinori Akiyama, Tamaki Kobayashi, Yukihiko Yamao, Takayuki Kikuchi, Riki Matsumoto, Takeharu Kunieda, et al., "Frontal fibers connecting the superior frontal gyrus to broca area: a corticocortical evoked potential study," *World neurosurgery*, vol. 107, pp. 239–248, 2017.
- [15] Natalie V Covington and Melissa C Duff, "Expanding the language network: Direct contributions from the hippocampus," *Trends in Cognitive Sciences*, vol. 20, no. 12, pp. 869–870, 2016.
- [16] Gabriel Möddel, Tara Lineweaver, Stephan U Schuele, Julia Reinholz, and Tobias Loddenkemper, "Atypical language lateralization in epilepsy patients," *Epilepsia*, vol. 50, no. 6, pp. 1505–1516, 2009.
- [17] Velia Cardin, Rebecca C Smittenaar, Eleni Orfanidou, Jerker Rönnerberg, Cheryl M Capek, Mary Rudner, and Bencie Woll, "Differential activity in heschl's gyrus between deaf and hearing individuals is due to auditory deprivation rather than language modality," *Neuroimage*, vol. 124, pp. 96–106, 2016.
- [18] Jeong-A Kim, Jeong-Won Jeong, Michael E Behen, Vinod K Pilli, Aimee Luat, Harry T Chugani, and Csaba Juhász, "Metabolic correlates of cognitive function in children with unilateral sturge-weber syndrome: Evidence for regional functional reorganization and crowding," *Human brain mapping*, vol. 39, no. 4, pp. 1596–1606, 2018.
- [19] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu, "A comprehensive survey on graph neural networks," *arXiv preprint arXiv:1901.00596*, 2019.