



Spatiotemporal Incremental Mechanics Modeling of Facial Tissue Change

Nathan Lampen¹, Daeseung Kim²(✉), Xuanang Xu¹, Xi Fang¹,
Jungwook Lee¹, Tianshu Kuang², Hannah H. Deng²,
Michael A. K. Liebschner³, James J. Xia², Jaime Gateno²,
and Pingkun Yan¹(✉)

¹ Department of Biomedical Engineering and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
yanp2@rpi.edu

² Department of Oral and Maxillofacial Surgery, Houston Methodist Research Institute, Houston, TX 77030, USA
dkim@houstonmethodist.org

³ Department of Neurosurgery, Baylor College of Medicine, Houston, TX 77030, USA

Abstract. Accurate surgical planning for orthognathic surgical procedures requires biomechanical simulation of facial soft tissue changes. Simulations must be performed quickly and accurately to be useful in a clinical pipeline, and surgeons may try several iterations before arriving at an optimal surgical plan. The finite element method (FEM) is commonly used to perform biomechanical simulations. Previous studies divided FEM simulations into incremental steps to improve convergence and model accuracy. While incremental simulations are more realistic, they greatly increase FEM simulation time, preventing integration into clinical use. In an attempt to make simulations faster, deep learning (DL) models have been developed to replace FEM for biomechanical simulations. However, previous DL models are not designed to utilize temporal information in incremental simulations. In this study, we propose Spatiotemporal Incremental Mechanics Modeling (SIMM), a deep learning method that performs spatiotemporally-aware incremental simulations for mechanical modeling of soft tissues. Our method uses both spatial and temporal information by combining a spatial feature extractor with a temporal aggregation mechanism. We trained our network using incremental FEM simulations of 18 subjects from our repository. We compared SIMM to spatial-only incremental and single-step simulation approaches. Our results suggest that adding spatiotemporal information may improve the accuracy of incremental simulations compared to methods that use only spatial information.

Keywords: Surgical Planning · Deep Learning · Spatiotemporal · Biomechanics

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43996-4_54.

1 Introduction

Orthognathic surgery corrects facial skeletal deformities that may cause functional and aesthetic impairments. In orthognathic surgical procedures, the jaws are cut into several bony segments and repositioned to desired locations to achieve an ideal alignment. Surgeons plan the movement of osteotomized bony segments before surgery to obtain the best surgical outcome. While surgeons do not operate on the soft tissue of the face during the procedure, it is passively moved by the underlying bone, causing a change in facial appearance. To visualize the proposed outcome, simulations of the planned procedure may be performed to predict the final facial tissue appearance. Current simulation techniques use the finite element method (FEM) to estimate the change of the facial tissue caused by the movement of the bony segments [2, 5, 6, 8, 9]. Subject-specific FEM models of the facial tissue are created from preoperative Computed Tomography (CT) imaging. The planned bony displacement is fed to the model as an input boundary condition. Previous studies divided the input bony displacement into smaller, incremental steps to improve model accuracy [7]. While incremental FEM simulations are very accurate, they require significant computational time, sometimes approaching 30 min to perform a single simulation [7]. This restricts the utility of FEM since surgeons may try several iterations before arriving at an optimal surgical plan, and simulations must be performed quickly to be useful in a clinical pipeline.

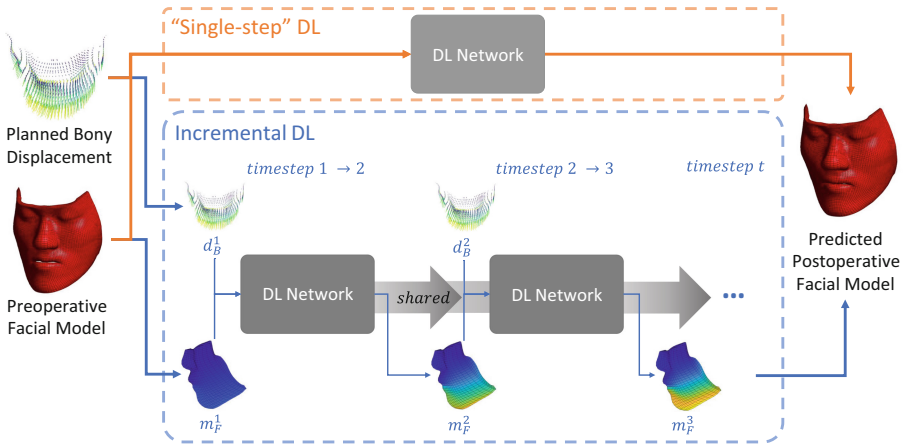


Fig. 1. An overview of the difference between the “single-step” DL method and the incremental DL method, which replicates incremental FEM simulations.

Previous studies have implemented deep learning (DL) methods to shorten soft tissue simulations. Physics-informed neural networks (PINNs) learn the constitutive model of soft tissues, however, these models do not generalize well to new settings [1, 11, 15]. Other studies learned a general biomechanical model by training a UNet-style network on FEM simulations, but these methods are

limited to grid-like structures and cannot represent irregular meshes used in FEM [12, 14]. PointNet++ has also been used to learn biomechanical deformation based on point cloud data representing nodes within a mesh [10]. While PointNet++ can represent nodes in an irregular mesh, it does not capture the connections (edges) between nodes. PhysGNN utilizes edge information by learning biomechanical deformation based on graphs created from FE meshes [16]. However, the main limitation of all the aforementioned DL methods is that they perform “single-step” simulations (Fig. 1), which are not ideal for modeling non-linear materials such as facial soft tissue, especially when large deformation ($> 1\text{mm}$) is involved [7]. Incremental simulations help deal with the non-linearity of soft tissue by simulating several transitional states, providing a more realistic and traceable final postoperative prediction.

A simplistic approach to performing incremental simulations using DL is to break a large simulation into smaller steps and perform them in sequential order, but independently. However, this does not allow the model to correct errors, causing them to accumulate over incremental steps. One previous work implemented a data augmentation strategy to minimize network error in incremental simulations of soft tissue [17]. However, the network used was not designed to utilize temporal information or capture temporal trends across incremental steps, such as deformation that follows a general trajectory, which can be utilized to improve incremental predictions. While another previous work incorporated long short-term memory into a convolutional neural network (CNN) to learn across multiple timesteps, the use of a CNN means the network is limited to representing regular grids without integration of edge information [4]. Therefore, a method is needed which can perform incremental simulations using both spatial and temporal information while accepting irregular mesh geometry in the form of graphs. There are several technical challenges to combining spatial and temporal information in a deep learning network. First, it can be challenging to capture spatiotemporal features that effectively represent the data, especially in cases where the objects change shape over time. Second, spatiotemporal networks are often computationally complex and prone to overfitting, which limits the model’s clinical utility. For this reason, an explainable spatiotemporal network, i.e. one that learns temporal trends from already extracted spatial features, while minimizing computational complexity is needed.

In this study, we hypothesize that utilizing both spatial and temporal information for incremental simulations can improve facial prediction accuracy over DL networks that perform single-step simulations or incremental simulations while only considering spatial information. Therefore, the purpose of this study is to introduce a spatiotemporal deep learning approach for incremental biomechanics simulation of soft tissue deformation. We designed a network that learns spatial features from incremental simulations and aggregates them across multiple incremental simulations to establish sequential continuity. The contributions of this work are (1) a method for combining spatial and temporal learning in an incremental manner (2) a method for performing incremental simulations while preserving knowledge from previous increments. Our proposed method successfully implements spatiotemporal learning by observing temporal trends from spatial features while minimizing computational complexity.

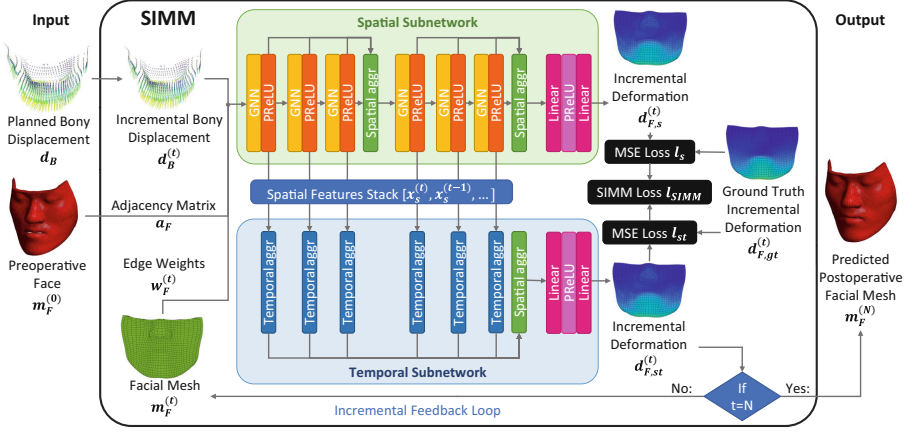


Fig. 2. The proposed framework of Spatiotemporal Incremental Mechanics Modeling (SIMM) for incremental mechanical modeling of facial tissue deformation.

2 Method

The Spatiotemporal Incremental Mechanics Modeling (SIMM) method predicts the incremental deformation of a three-dimensional (3D) soft tissue mesh based on incrementally planned input displacement. The SIMM network consists of two main operations: 1) spatial feature encoding and 2) temporal aggregation of spatial features (Fig. 2). The SIMM method is designed to first encode spatial features using a graph neural network optimized for spatial tasks, then observe how these features change over time using a temporal aggregation mechanism. SIMM is designed to be explainable and efficient by using the already extracted spatial features to observe temporal trends.

2.1 Data Representation

The SIMM method uses an incremental simulation scheme, shown in Fig. 2. SIMM is designed to accept input data in graph format that represents volumetric meshes used in FEM. In the input graphs, vertices correspond to nodes in an FE mesh and the edges correspond to the connections between nodes within an element [13]. For this reason, we will refer to our input data as “meshes” consisting of “nodes” and “edges” in the rest of this manuscript. In the SIMM method, the planned bony displacement d_B for a surgical procedure is broken into smaller steps and applied incrementally. The incremental planned bony displacement $d_B^{(t)}$ and geometric information from an input facial mesh $m_F^{(0)}$ are fed to the spatial sub-network to predict the incremental facial deformation $d_{F,s}^{(t)}$. The geometric information from the input facial mesh consists of an adjacency matrix $a_F^{(t)}$ and a set of edge weights $w_F^{(t)}$. The adjacency matrix is a binary matrix describing the connections between nodes. The edge weights are calculated as the inverse

of the Euclidean distance between nodes, providing the network with spatial information. Spatial features x_s are extracted in each GNN layer and aggregated using jumping-knowledge connections [18]. In the temporal sub-network, the stack of encoded spatial features $[x_s^{(t)}, x_s^{(t-1)}, \dots]$ are aggregated across several incremental steps, resulting in a separate spatiotemporal prediction of the incremental facial tissue deformation $d_{F,st}^{(t)}$. Both the spatial-only and spatiotemporal predictions are compared to the ground truth FEM predicted incremental facial deformation $d_{F,gt}^{(t)}$. The spatiotemporally predicted facial deformation is then used to update the facial mesh for the subsequent timestep. After all incremental steps are applied (at $t = N$), the final predicted deformation is used to calculate the predicted postoperative facial mesh m_F^N .

2.2 Network

Spatial sub-network: We used a PhysGNN network to learn the spatial features [16]. PhysGNN consists of six GNN layers that extract features in an increasing neighborhood around a given node. PhysGNN also uses jumping-knowledge connections which aggregate information from the first three and last three GNN layers. The jumping-knowledge connections use an aggregation function to fuse features across the GNN layers using the equation:

$$y_s = f_{agg}(x^{(0)}, \dots, x^{(l)}) \quad (1)$$

where y_s is the spatially aggregated features, and $x^{(l)}$ are the features from layers 0 to l . f_{agg} can be any form of aggregation function, such as concatenation. In this work, we use a long-short-term memory (LSTM) attention aggregation mechanism to achieve optimal results, which has been demonstrated in our ablation study (Sect. 3.4) [3, 18].

Temporal sub-network: To capture temporal trends, we implemented additional aggregation layers to process features across a sequence of incremental timepoints. Instead of aggregating features across several spatial “neighborhoods” as seen in Eq. (1), these additional aggregation layers aggregate features from a given spatial neighborhood across several sequential timepoints (t), as seen in Eq. (2):

$$y_t = f_{agg}(x^{(t)}, \dots, x^{(t-N)}) \quad (2)$$

where y_t is the temporally aggregated features and $x^{(t)}$ are the features to be aggregated across timepoints (t) to $(t - N)$. We hypothesized that including more than one previous timepoint in the aggregation layer may improve the network performance (see Sect. 3.4). We used the same LSTM-attention aggregation mechanism as used in the PhysGNN spatial sub-network.

Training Strategy: The SIMM network produces two predictions of the incremental tissue deformation, one from the spatial sub-network and the other after spatiotemporal aggregation (Fig. 2). We used incremental FEM simulations as ground truth while training our network. The loss was calculated as the mean-squared error between the predicted deformation and the ground truth FEM deformation. We use the following equation to calculate the loss of the network:

$$l_{SIMM} = l_s + l_{st} \quad (3)$$

where l_{SIMM} is the total loss of the SIMM network, l_s is the loss of the spatial sub-network predicted deformation $d_{F,s}^{(t)}$, and l_{st} is the loss of the spatiotemporal predicted deformation $d_{F,st}^{(t)}$. The combined loss was used to train the network to ensure the spatial sub-network still learns to adequately encode the spatial features.

3 Experiments and Results

3.1 Dataset

The SIMM method was evaluated on a dataset of incremental FEM simulations from 18 subjects using a leave-one-out cross-validation (LOOCV). The subjects were chosen from our digital archive of patients who underwent double-jaw orthognathic surgery (IRB# PRO00008890). FEM meshes were generated from CT scans and FE simulations were performed using an existing clinical pipeline [7]. Boundary conditions including bony movements and sliding effect were applied to the FEM mesh. FEM simulation was performed to simulate the facial change with at least 200 incremental steps to have enough training data. The total number of incremental steps varied, depending on the stability of each incremental simulation step. Meshes for all subjects consisted of 3960 nodes and 2784 elements. The incremental bony displacement information was fed to the network as an input feature vector for each node. For facial nodes assumed to move together with the underlying moving bony surface, the input displacement vector was equal to the bony surface displacement $d_B^{(t)}$, which describes the boundary condition between the moving bony structures and the facial soft tissue. For facial nodes not on the surface of a moving bony segment, or that are on the surface of a fixed bony segment, the input displacement vector was initialized as all zeros. To increase the number of simulations for training, we further split the simulation for a given subject into several “sub-simulations” consisting of a subset of the incremental timesteps. The timesteps for each sub-simulation were determined by using a maximum deformation threshold, d_{max} . For example, starting from timepoint 0, the first incremental timepoint with a maximum deformation exceeding the threshold would be included in the sub-simulation (Fig. S1). The starting point of each sub-simulation was unique, however, all sub-simulations ended on the postoperative face as the last incremental step, allowing for a direct comparison between all sub-simulations on the postoperative mesh for a given subject. We used a maximum deformation threshold of 0.5mm when splitting each subject’s simulation into sub-simulations, which was chosen based on an ablation study (see Sect. 3.4). The number of sub-simulations for each subject can be found in Tab. S1.

Table 1. Mean Euclidean distance error, mean simulation time, and model size comparison between SIMM, Incremental, and Single-step deep learning methods.

Method	Single-Step	Incremental	SIMM
Mean Error \pm StdDev [mm]	0.44 ± 0.14^a	0.47 ± 0.19^a	0.42 ± 0.13^b
Mean Simulation Time [s]	0.08	1.14	1.36
Model Size [MB]	2052	2774	3322
# of Parameters [M]	0.9	0.9	3.3

* Values with different superscript letters are significantly different ($p < 0.05$)

3.2 Implementation Details and Evaluation Metrics

All networks were trained in pytorch using the Adam optimizer with an initial learning rate of $5e-3$ for 100 epochs on an NVIDIA Tesla V100. We trained all networks in leave-one-out cross-validation across all 18 subjects. The loss was calculated as the mean squared error between the network-predicted and FEM-predicted deformation vectors for each node within the facial mesh. The final accuracy was calculated as the Euclidean distance between the network-predicted and FEM-predicted node coordinates on the final postoperative face. The distribution of the Euclidean distances was not found to be normal using the Kolmogorov-Smirnov test, so statistical significance between methods was tested using the Wilcoxon signed-rank test and a p -value of 0.05.

3.3 Comparison with Baselines

We compared our SIMM method with two baseline methods: 1) a single-step method, and 2) an incremental method. For the single-step method, we trained a PhysGNN network to predict the total facial deformation in a single step. The same sub-simulations used in the incremental and SIMM methods (see Sect. 3.1) were used to train the single-step network, however, all intermediate timepoints between the first and final timepoints were removed. For the incremental method, we used PhysGNN to perform incremental simulations. The incremental method used the prediction from timepoint $t-1$ to update the edge weights in timepoint t , similar to the feedback loop in SIMM (Fig. 2). However, no temporal aggregation or separate spatiotemporal prediction was used.

Our SIMM method achieved a mean error of 0.42 mm on all subjects (Table 1) with subject-specific mean error between 0.23 and 0.77 mm (Tab. S1). In comparison, the single-step method achieved a mean error of 0.44 mm on all subjects (Table 1) with subject-specific mean error between 0.30 and 0.91 mm (Tab. S1). The incremental method achieved a mean error of 0.47 mm on all subjects (Table 1) with subject-specific mean error between 0.25 and 1.00 mm (Tab. S1). Statistical analysis showed SIMM performed significantly better than the single-step and incremental methods, while the single-step and incremental methods were not significantly different.

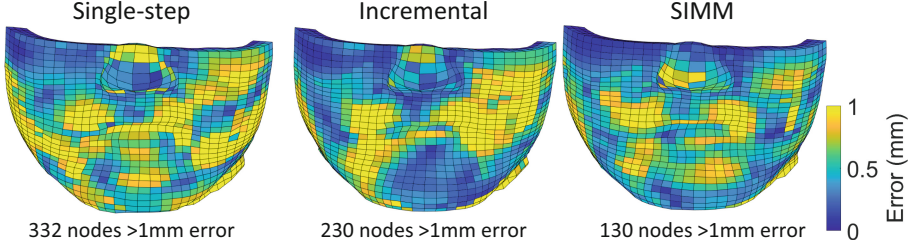


Fig. 3. Error of a) the single-step method, b) incremental method, and c) SIMM method for subject 9

3.4 Ablation Studies

We also performed ablation studies to investigate the effects of several key components in the SIMM method. First, when splitting subject simulations into sub-simulations, we varied the maximum deformation threshold d_{max} to 1.0, 0.5, and 0.1 mm. We found a d_{max} of 0.5 achieved the best performance in the incremental method, although the best d_{max} may change for different cases (Tab. S2). Second, we investigated the effect of the type of aggregation mechanism used in the PhysGNN network of the incremental method by replacing the LSTM-attention aggregation mechanism with a concatenation aggregation mechanism. The mean error increased to 1.25 mm when using a concatenation aggregation. Third, we tried increasing the number of previous timepoints in the spatial features stack to be used in the temporal aggregation layers from 1 previous timepoint to 5 previous timepoints. We hypothesized that including multiple previous timepoint in the aggregation layer may improve performance. The mean error increased to 0.44 mm when using 5 previous timepoints.

4 Discussions and Conclusions

The SIMM method achieved a lower mean error than the single-step and incremental methods, as seen in the quantitative results (Table 1). The results of the incremental method suggest the network accumulates errors across incremental steps, as seen in a plot of the predicted deformation vectors over time (Fig. S2). Figure 3 shows an example of the improvement of SIMM over the single-step and incremental methods. The results of the ablation study showed concatenation aggregation did not perform as well as LSTM-attention aggregation, which is consistent with other studies that investigated aggregation methods in GNNs [3, 18]. Our ablation study also demonstrated that increasing the number of previous timepoints in the temporal feature aggregation did not improve network performance. We think this result is likely caused by over-smoothing the temporal information when aggregating features from many previous timepoints.

One limitation of our SIMM method is that it requires incremental FEM simulations to be trained. This greatly increases training requirements over single-

step methods, which can feasibly be trained only on preoperative and postoperative facial data without performing incremental FEM simulations. However, this is true of any incremental simulation method. Once SIMM is trained, the inference simulation time is considerably faster than FEM, which can take several minutes to perform an incremental simulation. The SIMM method could be easily extended to biomechanical simulations of many other types of soft tissues and clinical applications.

In conclusion, we successfully created a spatiotemporal incremental mechanics modeling (SIMM) method for simulating incremental facial deformation for surgical planning. The SIMM method shows potential to improve simulation accuracy over methods based on spatial information alone, suggesting the importance of spatiotemporal learning for incremental simulations.

Acknowledgements. This work was partially supported by NIH awards R01 DE022676, R01 DE027251, and R01 DE021863.

References

1. Buoso, S., Joyce, T., Kozerke, S.: Personalising left-ventricular biophysical models of the heart using parametric physics-informed neural networks. *Med. Image Anal.* **71** (2021)
2. Chabanas, M., Luboz, V., Payan, Y.: Patient specific finite element model of the face soft tissues for computer-assisted maxillofacial surgery. *Med. Image Anal.* **7**(2), 131–151 (2003)
3. Hamilton, W.L., Ying, R., Leskovec, J.: Inductive representation learning on large graphs. In: *Advances in Neural Information Processing Systems 2017-December*, pp. 1025–1035 (2017)
4. Karami, M., Lombaert, H., Rivest-Hénault, D.: Real-time simulation of viscoelastic tissue behavior with physics-guided deep learning. *Comput. Med. Imaging Graph.* **104**, 102165 (2023)
5. Kim, D., et al.: A clinically validated prediction method for facial soft-tissue changes following double-jaw surgery. *Med. Phys.* **44**(8), 4252–4261 (2017)
6. Kim, D., et al.: A new approach of predicting facial changes following orthognathic surgery using realistic lip sliding effect. In: *Medical Image Computing and Computer-Assisted Intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 11768, pp. 336–344 (2019)
7. Kim, D., et al.: A novel incremental simulation of facial changes following orthognathic surgery using FEM with realistic lip sliding effect. *Med. Image Anal.* **72**, 102095 (2021)
8. Knoops, P.G., et al.: Three-dimensional soft tissue prediction in orthognathic surgery: a clinical comparison of Dolphin, ProPlan CMF, and probabilistic finite element modelling. *Int. J. Oral Maxillofac. Surgery* **48**(4), 511–518 (2019)
9. Knoops, P.G., et al.: A novel soft tissue prediction methodology for orthognathic surgery based on probabilistic finite element modelling. *PLOS ONE* **13**(5), e0197209 (2018)
10. Lampen, N., et al.: Deep learning for biomechanical modeling of facial tissue deformation in orthognathic surgical planning. *Int. J. Comput. Assist. Radiol. Surgery* **17**(5), 945–952 (2022)

11. Liu, M., Liang, L., Sun, W.: A generic physics-informed neural network-based constitutive model for soft biological tissues. *Comput. Methods Appl. Mech. Eng.* **372**, 113402 (2020)
12. Mendizabal, A., Márquez-Neila, P., Cotin, S.: Simulation of hyperelastic materials in real-time using deep learning. *Med. Image Anal.* **59**, 101569 (2020)
13. Pfaff, T., Fortunato, M., Sanchez-Gonzalez, A., Battaglia, P.: Learning mesh-based simulation with graph networks. In: *International Conference on Learning Representations* (2021)
14. Pfeiffer, M., Riediger, C., Weitz, J., Speidel, S.: Learning soft tissue behavior of organs for surgical navigation with convolutional neural networks. *Int. J. Comput. Assist. Radiol. Surgery* **14**(7), 1147–1155 (2019)
15. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving non-linear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019)
16. Salehi, Y., Giannacopoulos, D.: PhysGNN: a physics-driven graph neural network based model for predicting soft tissue deformation in image-guided neurosurgery. *arXiv preprint [arXiv:2109.04352](https://arxiv.org/abs/2109.04352)* (2021)
17. Wu, J.Y., Munawar, A., Unberath, M., Kazanzides, P.: Learning Soft-Tissue Simulation from Models and Observation. In: *2021 International Symposium on Medical Robotics, ISMR 2021* (2021)
18. Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K.I., Jegelka, S.: representation learning on graphs with jumping knowledge networks. In: *35th International Conference on Machine Learning, ICML 2018*, vol. 12, pp. 8676–8685 (2018)