WWW2023 Rebuttal #2442

Title: Predicting the Silent Majority on Graphs: Knowledge Transferable Graph Neural Network
December 21, 2022

1 Generalization Ability of GCN vs. Our Approach

To demonstrate the necessity of eliminating domain difference for the message passing mechanism of GNNs under our problem setting, we prove that GCN Welling and Kipf (2016) will have severely limited generalization ability if the domain difference between vocal nodes and silent nodes are ignored. However, our approach (Domain Adapted Knowledge Transfer) does not have such limitations and has robust generalization ability.

1.1 Theorem

In this section, we demonstrate that **our approach has stronger generalization ability on silent node classification over VS-Graphs than existing GCN methods**. And we quantify the generalization ability of GNNs by Complexity Measure (Consistency of Representations) Natekar and Sharma (2020); Du et al. (2022), where **higher complexity measure means lower generalization ability**. Then we give the following theorem under the assumptions in Sec. 2.2 to quantify the generalization ability of GCN Welling and Kipf (2016) and our KTGNN.

Theorem 1 : Unstable Generalization Ability of GCN. *For the binary silent node classification problem and the input graph satisfies the above assumptions, the complexity measure (Consistency of Representations) of GCN, denoted as* Γ*, has the following property* $\Gamma \propto \frac{1}{||\mathbf{W}\cdot(\alpha\cdot\overrightarrow{\Delta\mu}^v+\beta\cdot\overrightarrow{\Delta\mu}^s)||}$*, where* $\mathbf{W}$ *is the transformation matrix,* $\alpha = p_0^v + p_1^v - p_v$*,* $\beta = p_0^s + p_1^s - p_s$*,* $\overrightarrow{\Delta\mu}^v = \boldsymbol{\mu}_0^v - \boldsymbol{\mu}_1^v$ *and* $\overrightarrow{\Delta\mu}^s = \boldsymbol{\mu}_0^s - \boldsymbol{\mu}_1^s$*.* Γ *is strongly correlated to the directions of* $\overrightarrow{\Delta\mu}^s$ *and* $\overrightarrow{\Delta\mu}^v$*), which causes unstable generalization ability (even loses generalization ability) on the VS-Graphs.*

Theorem 2 : Robust Generalization Ability of KTGNN. *For the binary silent node classification problem and the input graph satisfies the above assumptions, the complexity measure (Consistency of Representations) of KTGNN* $\Gamma' \propto \frac{1}{||\mathbf{W}\cdot\gamma\cdot\overrightarrow{\Delta\mu}^s||}$*, where* $\mathbf{W}$ *is the transformation matrix,* $\gamma = p_0^v + p_1^v + p_0^s + p_1^s - 1$ *and* $\overrightarrow{\Delta\mu}^s = \boldsymbol{\mu}_0^s - \boldsymbol{\mu}_1^s$*.* Γ' *is uncorrelated to the directions of* $\overrightarrow{\Delta\mu}^v$*, and is invariant to the distribution difference between vocal nodes and silent nodes. Thus KTGNN has robust generalization ability that is insensitive to the domain-shift.*

The proof of Theorem 1 and Theorem 2 can be found at Sec. 2.3 and Sec. 2.4.

1.2 Insights of Theorems

Based on Theorem 1 and Theorem 2, we can conclude that **our approach has better generalization ability on silent node classification over VS-Graphs than existing GCN methods**.

Based on Theorem 1, we know that $\Gamma \propto \frac{1}{||\mathbf{W}\cdot(\alpha\cdot\overrightarrow{\Delta\mu}^v+\beta\cdot\overrightarrow{\Delta\mu}^s)||}$. Therefore the generalization ability of GCN-based model is **strongly correlated to the domain difference between vocal nodes and silent nodes** (the different directions of $\overrightarrow{\Delta\mu}^s$ and $\overrightarrow{\Delta\mu}^v$) on the VS-Graphs. When the directions of $\overrightarrow{\Delta\mu}^s$ and $\overrightarrow{\Delta\mu}^v$ are quite different (e.g., the component of them will be counteracted when the angle between the two vectors is larger than ninety degree), the generalization ability of GCN is quite unstable and my even lose generalization completely (Γ of GCN will converge to $+\infty$ when $\overrightarrow{\Delta\mu}^s = -\frac{\alpha}{\beta}\cdot\overrightarrow{\Delta\mu}^v$, and thus GCN will lose generalization ability). Theorem 1 indicates that GCN has unstable generalization ability on the silent node classification task.

Based on Theorem 2, we know that $\Gamma' \propto \frac{1}{||\mathbf{W}\cdot\gamma\cdot\overrightarrow{\Delta\mu}^s||}$. Therefore, the generalization ability of our approach is **uncorrelated to the directions of** $\overrightarrow{\Delta\mu}^v$, and eliminates the negative effects of the domain-shift between vocal nodes and silent nodes that may cause loss of generalization ability. Theorem 2 indicates that our approach has robust generalization ability on the silent node classification task.

Finally we can get the conclusion that our approach has more robust generalization ability on silent node classification over VS-Graphs than existing GCN methods.

2 Proof

2.1 Backgrounds: Complexity Measure

We aim at proving the generalization ability of our method over existing GCN-based models. To represent the generalization ability of GCN-based models, we introduce Complexity Measure Neyshabur et al. (2017), which is the current mainstream method to measure the generalization ability of the model. And we choose Consistency of Representations (Champion solution of the NeurIPS2020 Competition on Predicting Generalization in Deep Learning) Natekar and Sharma (2020); Du et al. (2022). And the complexity measure of the model, denoted as Γ, is:

$$\Gamma = \frac{1}{|\mathcal{C}|} \sum_{i=1}^{|\mathcal{C}|} \max_{i \neq j} \frac{\mathcal{S}_i + \mathcal{S}_j}{\mathcal{M}_{i,j}}, \tag{1}$$

where $\mathcal{C}$ is the set of classes, $\mathcal{C}_i, \mathcal{C}_j \in \mathcal{C}$ are two different classes (e.g., $\mathcal{C}_0 = \{v_k | y_k = 0\}$), $\mathcal{S}_i = (\mathbb{E}_{v_k \sim \mathcal{C}_i}(|h_k - \mu_{\mathcal{C}_i}|^p))^{\frac{1}{p}}$ is the intra-class variance of class $\mathcal{C}_i$, $\mathcal{M}_{i,j} = ||\mu_{\mathcal{C}_i} - \mu_{\mathcal{C}_j}||_p$ is the inter-class variance between $\mathcal{C}_i$ and $\mathcal{C}_j$, h_k is the representations of nodes v_k learned by the model and $\mu_{\mathcal{C}_i} = \mathbb{E}_{v_k \sim \mathcal{C}_i}(h_k)$. And **higher complexity measure means lower generalization ability**.

2.2 Hypothesis of Theorems

To simplify the analysis, we further formalize the problem with the following hypothesis:

1. We consider a **silent node classification** problem on the VS-Graph, with the **binary class set** $Y = \{0, 1\}$, and the **node population set** $\mathcal{O} = \{vocal, silent\}$.
2. We assume that the feature distribution conditioned on the class label and node population (vocal/silent node) follows the **Gaussian Distribution**. Specifically, we assume that:
   - $P(X|Y = 0, \mathcal{O} = vocal) \sim \mathbf{N}(\boldsymbol{\mu}_0^v, \boldsymbol{\sigma}_0^{v2})$; $\quad P(X|Y = 1, \mathcal{O} = vocal) \sim \mathbf{N}(\boldsymbol{\mu}_1^v, \boldsymbol{\sigma}_1^{v2})$
   - $P(X|Y = 0, \mathcal{O} = silent) \sim \mathbf{N}(\boldsymbol{\mu}_0^s, \boldsymbol{\sigma}_0^{s2})$; $\quad P(X|Y = 1, \mathcal{O} = silent) \sim \mathbf{N}(\boldsymbol{\mu}_1^s, \boldsymbol{\sigma}_1^{s2})$

   where $\boldsymbol{\mu}_0^v \neq \boldsymbol{\mu}_1^v \neq \boldsymbol{\mu}_0^s \neq \boldsymbol{\mu}_1^s$ are mean vectors, $\boldsymbol{\sigma}_0^v \neq \boldsymbol{\sigma}_1^v \neq \boldsymbol{\sigma}_0^s \neq \boldsymbol{\sigma}_1^s$ are variance vectors.
3. Given any silent node $v_i \in V^{silent}$ and its class label y_i, we assume the **neighborhood distribution** of its neighbor's label and population is known. Specifically, the probability of its neighbor $v_j \in \mathcal{N}(v_i)$ belonging to V^{vocal} (or V^{silent}) is p_v (or p_s), and $p_v + p_s = 1$. Then the joint distribution of neighbor's label and population:
   - $P(\mathcal{O}_j = vocal, Y_j = 0 | v_j \in \mathcal{N}(v_i), Y_i = 0) = p_0^v$
   - $P(\mathcal{O}_j = vocal, Y_j = 1 | v_j \in \mathcal{N}(v_i), Y_i = 0) = p_v - p_0^v$
   - $P(\mathcal{O}_j = silent, Y_j = 0 | v_j \in \mathcal{N}(v_i), Y_i = 0) = p_0^s$
   - $P(\mathcal{O}_j = silent, Y_j = 1 | v_j \in \mathcal{N}(v_i), Y_i = 0) = p_s - p_0^s$
   - Similarly, given $Y_i = 1$, the corresponding joint distribution changes to $p_1^v, p_v - p_1^v, p_1^s, p_s - p_1^s$.

2.3 Proof of Theorem 1

We analyze the generalization ability of GCN by derive the complexity measure (Consistency of Representations Natekar and Sharma (2020); Du et al. (2022)).Then we consider one GCN layer with mean aggregator as follows:

$$h_i = \sum_{j \in \mathcal{N}(v_i)} \frac{1}{d_i} \cdot X_j \cdot \mathbf{W} = \left(\frac{1}{d_i} \cdot \sum_{j \in \mathcal{N}(v_i)} X_j \right) \cdot \mathbf{W} \tag{2}$$

To simplify the analysis, we omit the activation function. h_i denotes the representation of v_i, X_i is the raw feature of v_i, $W^{(l)}$ is the transformation matrix, d_i is the degree of v_i and $\mathcal{N}(v_i)$ indicates the neighbors of v_i. Based on Eq. 1, we need to calculate the intra-class variance $\{\mathcal{S}_0, \mathcal{S}_1\}$ and inter-class variance $\mathcal{M}_{0,1}$ for the representations of silent nodes. The class center of silent nodes in $\mathcal{C}_0$ class, denoted by $\boldsymbol{\mu}^s_{H_0}$, is:

$$\begin{aligned}\boldsymbol{\mu}^s_{H_0} &= \mathbb{E}_{v_i \in V^{silent}, Y_i=0}\left[\frac{1}{d_i}\cdot \mathbf{W}\cdot\Big(\sum_{v_j\in\mathcal{N}(v_i)\cap V^{vocal}} X_j + \sum_{v_k\in\mathcal{N}(v_i)\cap V^{silent}} X_k\Big)\right] \\ &= \mathbf{W}\cdot[p_0^v\cdot\boldsymbol{\mu}_0^v + (p_v - p_0^v)\cdot\boldsymbol{\mu}_1^v + p_0^s\cdot\boldsymbol{\mu}_0^s + (p_s - p_0^s)\cdot\boldsymbol{\mu}_1^s]\end{aligned} \tag{3}$$

Similarly, we have the class center of silent nodes in $\mathcal{C}_1$ class, denoted by $\boldsymbol{\mu}^s_{h_1}$:

$$\begin{aligned}\boldsymbol{\mu}^s_{H_1} &= \mathbb{E}_{v_i \in V^{silent}, Y_i=1}\left[\frac{1}{d_i}\cdot \mathbf{W}\cdot\Big(\sum_{v_j\in\mathcal{N}(v_i)\cap V^{vocal}} X_j + \sum_{v_k\in\mathcal{N}(v_i)\cap V^{silent}} X_k\Big)\right] \\ &= \mathbf{W}\cdot[(p_v - p_1^v)\cdot\boldsymbol{\mu}_0^v + p_1^v\cdot\boldsymbol{\mu}_1^v + (p_s - p_1^s)\cdot\boldsymbol{\mu}_0^s + p_1^s\cdot\boldsymbol{\mu}_1^s]\end{aligned} \tag{4}$$

Then we can calculate $\mathcal{M}_{0,1}$ based on Eq. 3 and Eq. 4:

$$\begin{aligned}\mathcal{M}_{0,1} &= \left|\left|\boldsymbol{\mu}^s_{H_0} - \boldsymbol{\mu}^s_{H_1}\right|\right| \\ &= ||\mathbf{W}\cdot[(p_0^v + p_1^v - p_v)\cdot(\boldsymbol{\mu}_0^v - \boldsymbol{\mu}_1^v) + (p_0^s + p_1^s - p_s)\cdot(\boldsymbol{\mu}_0^s - \boldsymbol{\mu}_1^s)]|| \\ &= \left|\left|\mathbf{W}\cdot(\alpha\cdot\overrightarrow{\Delta\mu}^v + \beta\cdot\overrightarrow{\Delta\mu}^s)\right|\right|\end{aligned} \tag{5}$$

where $\mathbf{W}$ is the transformation matrix of graph convolutional layer. To make the formula look more intuitive, we further simplify it by making $\alpha = p_0^v + p_1^v - p_v$, $\beta = p_0^s + p_1^s - p_s$, $\overrightarrow{\Delta\mu}^v = \boldsymbol{\mu}_0^v - \boldsymbol{\mu}_1^v$ and $\overrightarrow{\Delta\mu}^s = \boldsymbol{\mu}_0^s - \boldsymbol{\mu}_1^s$.

Then we view h_i and X_i as random variables and can further calculate the intra-class variance term $\mathcal{S}_0$ and $\mathcal{S}_1$:

$$\begin{aligned}{\mathcal{S}_0}^2 &= \mathbb{E}_{v_i\in V^{silent}, Y_i=0}\left[\left|\left|h_i - \boldsymbol{\mu}^s_{H_0}\right|\right|^2\right] \\ &= \mathbb{E}_{v_i\in V^{silent}, Y_i=0}\left[\left\langle h_i - \boldsymbol{\mu}^s_{H_0}, h_i - \boldsymbol{\mu}^s_{H_0}\right\rangle\right] \\ &= (p_0^v)^2\cdot\mathbb{E}_{v_i\in V^{vocal}, Y_i=0}\left[||\mathbf{W}(X_i - \boldsymbol{\mu}_0^v)||^2\right] + (p_v - p_0^v)^2\cdot\mathbb{E}_{v_i\in V^{vocal}, Y_i=1}\left[||\mathbf{W}(X_i - \boldsymbol{\mu}_1^v)||^2\right] \\ &\quad + (p_0^s)^2\cdot\mathbb{E}_{v_i\in V^{silent}, Y_i=0}\left[||\mathbf{W}(X_i - \boldsymbol{\mu}_0^s)||^2\right] + (p_s - p_0^s)^2\cdot\mathbb{E}_{v_i\in V^{silent}, Y_i=1}\left[||\mathbf{W}(X_i - \boldsymbol{\mu}_1^s)||^2\right]\end{aligned} \tag{6}$$

Based on the reparameterization trick proposed in Kingma and Welling (2013), a random variable X that follows Gaussian Distribution $\mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ can be converted to $\boldsymbol{\mu} + \epsilon\odot\boldsymbol{\sigma}$, where ϵ is a random variable and $\epsilon\sim\mathbf{N}(\mathbf{0},\mathbf{1})$. And then Eq 6 can be further simplified by the reparameterization trick:

$$\begin{aligned}{\mathcal{S}_0}^2 &= \mathbb{E}_{\epsilon\sim\mathbf{N}(\mathbf{0},\mathbf{1})}\Big[(p_0^v)^2\cdot||\mathbf{W}(\epsilon\odot\boldsymbol{\sigma}_0^v)||^2 + (p_v - p_0^v)^2\cdot||\mathbf{W}(\epsilon\odot\boldsymbol{\sigma}_1^v)||^2 \\ &\qquad + (p_0^s)^2\cdot||\mathbf{W}(\epsilon\odot\boldsymbol{\sigma}_0^s)||^2 + (p_s - p_0^s)^2\cdot||\mathbf{W}(\epsilon\odot\boldsymbol{\sigma}_1^s)||^2\Big] \\ &= \mathcal{F}_{\mathcal{S}_0}(\boldsymbol{\sigma}_0^v, \boldsymbol{\sigma}_1^v, \boldsymbol{\sigma}_0^s, \boldsymbol{\sigma}_1^s, p_0^v, p_0^s)\end{aligned} \tag{7}$$

Then we can conclude that the intra-class variance term is only related to the variance σ^2 of the four Gaussian Distribution in the second assumption (while unrelated to the expectation μ). And similarity we can get $\mathcal{S}_1$:

$$\begin{aligned}{\mathcal{S}_1}^2 &= \mathbb{E}_{\epsilon\sim\mathbf{N}(\mathbf{0},\mathbf{1})}\Big[(p_v - p_1^v)^2\cdot||\mathbf{W}(\epsilon\odot\boldsymbol{\sigma}_0^v)||^2 + (p_1^v)^2\cdot||\mathbf{W}(\epsilon\odot\boldsymbol{\sigma}_1^v)||^2 \\ &\qquad + (p_s - p_1^s)^2\cdot||\mathbf{W}(\epsilon\odot\boldsymbol{\sigma}_0^s)||^2 + (p_1^s)^2\cdot||\mathbf{W}(\epsilon\odot\boldsymbol{\sigma}_1^s)||^2\Big] \\ &= \mathcal{F}_{\mathcal{S}_1}(\boldsymbol{\sigma}_0^v, \boldsymbol{\sigma}_1^v, \boldsymbol{\sigma}_0^s, \boldsymbol{\sigma}_1^s, p_1^v, p_1^s)\end{aligned} \tag{8}$$

Finally we can get the complexity measure (Consistency of Representations) of GCN with Eq. 5∼8:

$$\begin{aligned}\Gamma &= \frac{\mathcal{S}_0 + \mathcal{S}_1}{\mathcal{M}_{0,1}} \\ &= \frac{\sqrt{\mathcal{F}_{\mathcal{S}_0}\left(\boldsymbol{\sigma}_0^v, \boldsymbol{\sigma}_1^v, \boldsymbol{\sigma}_0^s, \boldsymbol{\sigma}_1^s, p_0^v, p_0^s\right)} + \sqrt{\mathcal{F}_{\mathcal{S}_1}\left(\boldsymbol{\sigma}_0^v, \boldsymbol{\sigma}_1^v, \boldsymbol{\sigma}_0^s, \boldsymbol{\sigma}_1^s, p_1^v, p_1^s\right)}}{\left|\left|\mathbf{W} \cdot (\alpha \cdot \overrightarrow{\Delta\mu}^v + \beta \cdot \overrightarrow{\Delta\mu}^s)\right|\right|}\end{aligned} \tag{9}$$

when $\overrightarrow{\Delta\mu}^s = -\frac{\alpha}{\beta} \cdot \overrightarrow{\Delta\mu}^v$, $\mathcal{M}_{0,1}$ (the denominator of Eq 9) reaches 0 and $\Gamma \to +\infty$, thus GCN will lose generalization ability. Up to now, the Theorem 1 has been proved.

2.4 Proof of Theorem 2

Based on the design of Domain Adapted Message Passing (DAMP) module in our KTGNN, we alibrate the source nodes to the domain of target nodes before passing the messages from different domains to the target nodes. Considering a nonlinear mapping function $\Phi^{v \to s}$ (corresponding to the domain-shift calibration function in KTGNN), and for simplification, we consider a mean aggregator:

$$h_i' = \left(\frac{1}{d_i} \cdot \sum_{j \in \mathcal{N}(v_i)} \Phi(X_j)\right) \cdot \mathbf{W} \tag{10}$$

Then the class center of silent nodes for representation learned by KTGNN (denoted by ${\boldsymbol{\mu}_{H_0}^s}'$ and ${\boldsymbol{\mu}_{H_1}^s}'$) changes to:

$$\begin{aligned}{\boldsymbol{\mu}_{H_0}^s}' &= \mathbb{E}_{v_i \in V^{silent}, Y_i = 0}\left[\frac{1}{d_i} \cdot \mathbf{W} \cdot \Big(\sum_{v_j \in \mathcal{N}(v_i) \cap V^{vocal}} \Phi(X_j) + \sum_{v_k \in \mathcal{N}(v_i) \cap V^{silent}} X_k\Big)\right] \\ &= \mathbf{W} \cdot [(p_0^v + p_0^s) \cdot \boldsymbol{\mu}_0^s + (1 - p_0^v - p_0^s) \cdot \boldsymbol{\mu}_1^s]\end{aligned} \tag{11}$$

$$\begin{aligned}{\boldsymbol{\mu}_{H_1}^s}' &= \mathbb{E}_{v_i \in V^{silent}, Y_i = 1}\left[\frac{1}{d_i} \cdot \mathbf{W} \cdot \Big(\sum_{v_j \in \mathcal{N}(v_i) \cap V^{vocal}} \Phi(X_j) + \sum_{v_k \in \mathcal{N}(v_i) \cap V^{silent}} X_k\Big)\right] \\ &= \mathbf{W} \cdot [(1 - p_1^v - p_1^s) \cdot \boldsymbol{\mu}_0^s + (p_1^v + p_1^s) \cdot \boldsymbol{\mu}_1^s]\end{aligned} \tag{12}$$

And we can get the new inter-class variance:

$$\begin{aligned}{\mathcal{M}_{0,1}}' &= \left|\left|{\boldsymbol{\mu}_{H_0}^s}' - {\boldsymbol{\mu}_{H_1}^s}'\right|\right| \\ &= ||\mathbf{W} \cdot [(p_0^v + p_1^v - p_v + p_0^s + p_1^s - 1) \cdot (\boldsymbol{\mu}_0^s - \boldsymbol{\mu}_1^s)]|| \\ &= \left|\left|\mathbf{W} \cdot \gamma \cdot \overrightarrow{\Delta\mu}^s\right|\right|\end{aligned} \tag{13}$$

where $\gamma = p_0^v + p_1^v - p_v + p_0^s + p_1^s - 1$ is a constant number, and $\overrightarrow{\Delta\mu}^s = \boldsymbol{\mu}_0^s - \boldsymbol{\mu}_1^s$. Without changing the variance σ in KTGNN, the intra-class variance $\mathcal{S}_0$ and $\mathcal{S}_1$ remain the same as Eq. 7 and Eq. 8. Finally we can get the complexity measure (Consistency of Representations) of KTGNN:

$$\begin{aligned}\Gamma' &= \frac{\mathcal{S}_0 + \mathcal{S}_1}{{\mathcal{M}_{0,1}}'} \\ &= \frac{\sqrt{\mathcal{F}_{\mathcal{S}_0}\left(\boldsymbol{\sigma}_0^v, \boldsymbol{\sigma}_1^v, \boldsymbol{\sigma}_0^s, \boldsymbol{\sigma}_1^s, p_0^v, p_0^s\right)} + \sqrt{\mathcal{F}_{\mathcal{S}_1}\left(\boldsymbol{\sigma}_0^v, \boldsymbol{\sigma}_1^v, \boldsymbol{\sigma}_0^s, \boldsymbol{\sigma}_1^s, p_1^v, p_1^s\right)}}{\left|\left|\mathbf{W} \cdot \gamma \cdot \overrightarrow{\Delta\mu}^s\right|\right|}\end{aligned} \tag{14}$$

From Eq. 14, it is obvious that Γ' remain invariant to the domain difference between vocal nodes and silent nodes (reflected by the different directions between $\overrightarrow{\Delta\mu}^s$ and $\overrightarrow{\Delta\mu}^v$). Therefore our KTGNN breaks the limitations of GCN by domain-shift colibration function Φ, and have robust generalization ability on the VS-Graphs compared with GCN. This also motivates us to design our KTGNN to transfer knowledge across domains. Up to now, the Theorem 2 has been proved.

References

Lun Du, Xiaozhou Shi, Qiang Fu, Xiaojun Ma, Hengyu Liu, Shi Han, and Dongmei Zhang. 2022. GBK-GNN: Gated Bi-Kernel Graph Neural Networks for Modeling Both Homophily and Heterophily. In *Proceedings of the ACM Web Conference 2022*. 1550–1558.

Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

Parth Natekar and Manik Sharma. 2020. Representation based complexity measures for predicting generalization in deep learning. *arXiv preprint arXiv:2012.02775* (2020).

Behnam Neyshabur, Srinadh Bhojanapalli, David McAllester, and Nati Srebro. 2017. Exploring generalization in deep learning. *Advances in neural information processing systems* 30 (2017).

Max Welling and Thomas N Kipf. 2016. Semi-supervised classification with graph convolutional networks. In *J. International Conference on Learning Representations (ICLR 2017)*.