Hierarchy2vec–Representation Learning in Hierarchical Collaboration Networks for Team Performance Prediction

Sulyun Lee* and Kang Zhao‡

*Interdisciplinary Graduate Program in Informatics, University of Iowa

‡Department of Business Analytics, University of Iowa

*, †, ‡{sulyun-lee kang-zhao}@uiowa.edu

Abstract

In collaboration networks, synergies among team members play an important role in team performance. At the same time, many teams feature hierarchical structures, where some team members report to another with a higher level position. With members having different responsibilities or playing different roles in a team, collaboration patterns between team members organized in a hierarchy can have important implications for the team's performance. Focusing on the hierarchical nature of collaboration networks, this work is the first attempt to learn vector representations of teams based on individual team members' characteristics and how they are organized into a hierarchical team. The proposed hierarchy2vec model first learns node embeddings via hierarchically biased walks and then aggregates such node embeddings in a hierarchical way to generate the team's vector representations. Experiments on a real-world dataset for coaches in the National Football League (NFL) reveal that the proposed model can achieve better results in team performance predictions.

1 Introduction

Teamwork is about collaboration, where individuals in a team work together to reach the common goal with their own knowledge, skills, or expertise. Collaboration can boost work productivity and improve team performance by overcoming the limited skillset of a solo team [4, 8, 13]. For example, academic scholars with diverse disciplines, expertise, and background collaborate with each other to solve interdisciplinary research questions. Such collaborations can encourage intellectual stimulation by sharing the research perspectives from different disciplines or broaden their insights through sharing different cultural backgrounds [22], leading to high-quality research [1]. Beyond academic collaboration, companies interact and cooperate with others, such as connecting with suppliers, partners, or providers that enable access to various resources [25, 29, 36]. Such partnerships promote great synergy of firms that leads to a high performance [10]. Back to the individual level, collaborations can also be observed in teams for online gaming [20], or sports [5] where players in the same team collaborate to win games.

We can view these various collaborations from the network perspective because a better understanding of network patterns associated with better team performance is valuable for decision- or policy-making [11, 27, 33]. For example, a sports team can design a better lineup of players for winning games [23] or a company can find the best hire for a project team [7] for team success.

However, existing studies that predict team performance consider collaboration relationships among team members as homogeneous when learning individual and team representations. This assumption is not true in real-world teams since there are both vertical and horizontal relationships among team members [19]. In other words, hierarchical structures are widely observed in collaborations. For example, leaders in teams usually have more authority on decision-making and the responsibility of directing other team members [18]. Sometimes, a team can have more than one level of hierarchy. Since leaders play different roles in teams, their ties with other members they lead should be treated differently than ties among non-leading members.

Therefore, we propose an approach that captures the hierarchical relationships among team members while preserving both individual team members' characteristics and the hierarchical collaboration structures in a team. Specifically, the hierarchy2vec model first learns team embeddings that represent individual members in a team by hierarchy-aware walks and then aggregates individuals' embeddings in a hierarchical way into a team representation. Learned team embeddings are then used as feature vectors to predict team performance.

We evaluated the proposed model using a dataset of coaches from the National Football League (NFL). Each NFL team consists of multiple coaches with specific positions and is organized into a 3-levels hierarchy (head coaches, coordinators, and position coaches), making it an appropriate context to study team performance from the perspective of hierarchical collaboration networks. Results demonstrated that team embeddings learned by our model could better predict team performance than treating all team members and all collaboration ties the same.

2 Related Work

2.1 Collaboration network analysis

Many studies have viewed team formations and collaborations from the network perspective. Such collaboration networks can represent the inter-connectivity among interacting entities, thus enabling extensive analyses. For instance, some studies generated academic collaboration networks among scholars to predict the future academic collaborations [3, 6] and reveal the relationship between collaboration patterns and individual or team performance [1, 24, 40]. One study analyzed business collaboration networks to discover the impact of leadership distributions or divisions in a company's teams on the team performance [27]. Another research explored the determinants of efficient decision-making in companies by looking at the collaborative patterns found in collaboration networks [37].

Of particular interests are studies that focused on developing prediction models for team performance based on patterns in collaboration networks. Such studies used their own feature sets that are associated with team performance to design the prediction models. Those feature sets include dynamics of historical team performance, and collaborative patterns appeared in collaboration networks [2], average and maximum of individual feature values in a team that indicates the structural centrality and familiarity in networks [14].

However, these methods require manually engineering aggregated statistics of nodes or networks to represent the whole team. Individual team members are treated as equally important in the process, and dyadic connections between two team members are not considered.

2.2 Graph representation learning

2.2.1 Node embedding

In machine learning research, nodes in a network (a.k.a., graph) can be represented as low-dimensional vectors that preserve its network properties, such that similar nodes have similar representations [9, 39]. DeepWalk [30], node2vec [16], and struc2vec [31] are unsupervised node embedding approaches, which uses the Skip-gram model to learn node embeddings based on node sequences generated by random or biased walks in a graph [28].

Unlike unsupervised node embedding approaches, supervised node embedding approaches use end-to-end deep architectures that are trained to maximize or minimize a specific objective. For instance, Graph Convolutional Network (GCN) [21], GraphSAGE [17], Graph Attention Network [35] adopted the deep learning architectures and trained the node embeddings by aggregating the information gathered from the neighboring nodes and guiding the model training by the node labels.

Nevertheless, these node representation learning methods do not explicitly consider the heterogeneous ties that connect nodes at different levels of a hierarchy. While supervised approaches can potentially learn the different roles of different collaboration ties, doing so would require a large amount of training data, especially when the goal is to predict team performance based on node embeddings instead of predicting node labels directly.

2.2.2 Representation learning for collaboration networks

Few studies learned node representations in collaboration networks. One study focused only on collaborations as teams and learned the similar node embeddings if nodes have ever collaborated with each other in the past [38]. Another study focused on the magnitude of positive or negative influence among team members and optimized the node embeddings by projecting nodes with stronger collaborative influence into closer vector space [32]. The study by Gong et al. [15] is the only one that learns the representations of teams in collaboration networks for team performance prediction. They used the online gaming collaboration network and learned the team embedding by aggregating the individual team member information gathered from co-play gaming networks [34]. However, this study treated all team members the same while aggregating the team member information, and the hierarchical structure in teams has been ignored.

Thus, our work has two major contributions. *First*, we adopted a hierarchically biased random walk to learn node embeddings in hierarchical teams. *Second*, we hierarchically aggregated node embeddings to learn team embeddings for team performance prediction.

3 Methods

3.1 Overview

The aim of the proposed hierarchy2vec model is to learn a vector representation for a hierarchical team based on the collaboration network and predict the team's performance. In sum, hierarchy2vec consists of three steps:

(1) Construct a hierarchical collaboration network based on the hierarchical relationship among members of teams.

(2) Generate node features by concatenating nodes' individual feature and node embeddings learned via hierarchically biased walks on historical hierarchical collaboration networks.

(3) Aggregate node features in the same team in a hierarchical way and use an end-to-end learning architecture to learn team representations and predict team performance.

Figure 1 illustrates the architecture of the model.

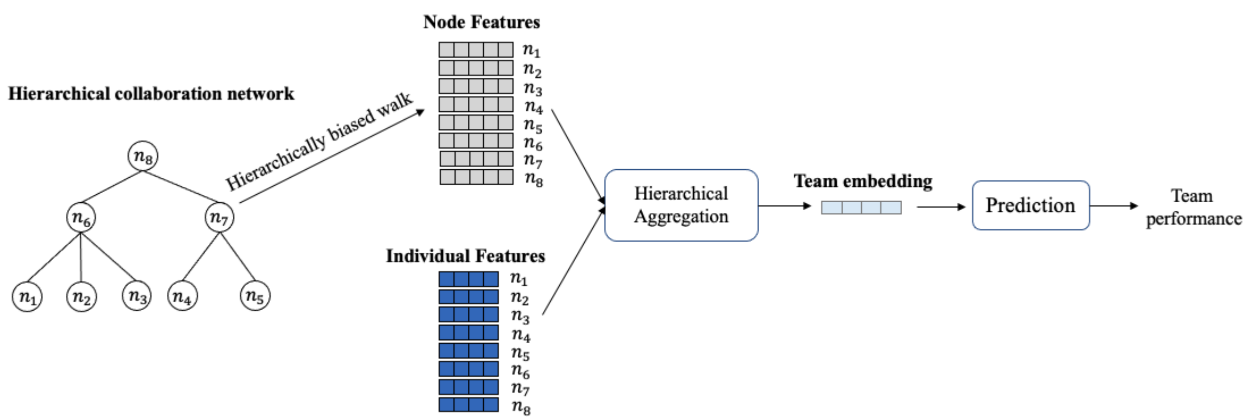

Figure 1: Overview of model architecture

3.2 Hierarchical collaboration networks

One input to the model is the cumulative collaboration network, which represents the hierarchical collaboration relationships among team members. In the network, nodes are individual team members, such as scholars in an academic team or players in a sports team. While the traditional way to represent relationships among members of the same team is to use a fully connected network among team members, a hierarchical collaboration network has three types of ties: (1) supervision ties represent "supervising" relationships, and point from a member at a higher level in the hierarchy to other members that she supervises (e.g., from a mentor to her apprentice); (2) reporting ties reflect "reporting to" or "working for" relationships. For each supervision tie, there is a reporting tie that reciprocates its direction (e.g., going from a student to a faculty advisor); (3) peer ties exist between colleagues working for the same supervisor, and such ties have no directions (e.g., between students working in the same lab).

Formally, a hierarchical collaboration network is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of nodes representing individual team members and $\mathcal{E}$ is a set of edges representing the hierarchical collaboration ties in the same team. There are three types of edges, $\{\mathcal{E}^{(\mathcal{S})}, \mathcal{E}^{(\mathcal{R})}, \mathcal{E}^{(\mathcal{P})}\} \in \mathcal{E}$, where $\mathcal{E}^{(\mathcal{S})}$ is a set of supervision ties, $\mathcal{E}^{(\mathcal{R})}$ is a set of reporting ties, and $\mathcal{E}^{(\mathcal{P})}$ is a set of peer ties.

3.3 Features

Node features include (1) Individual features that are relevant for team performance prediction, such as expertise, skills, or previous performance; (2) Collaboration features are learned from historical hierarchical collaboration networks. They are another important aspect that could have predictive power for team performance since team members' experience working for prestigious or high-performing leaders has a great impact on individuals' future career [12, 26].

Therefore, we extract nodes' collaboration features by learning from team members' previous collaboration with others. This process starts with constructing a cumulative collaboration network by combining hierarchical collaboration networks for teams assembled during the training period. Figure 2 provides a simple example, where two teams' hierarchical networks are combined.

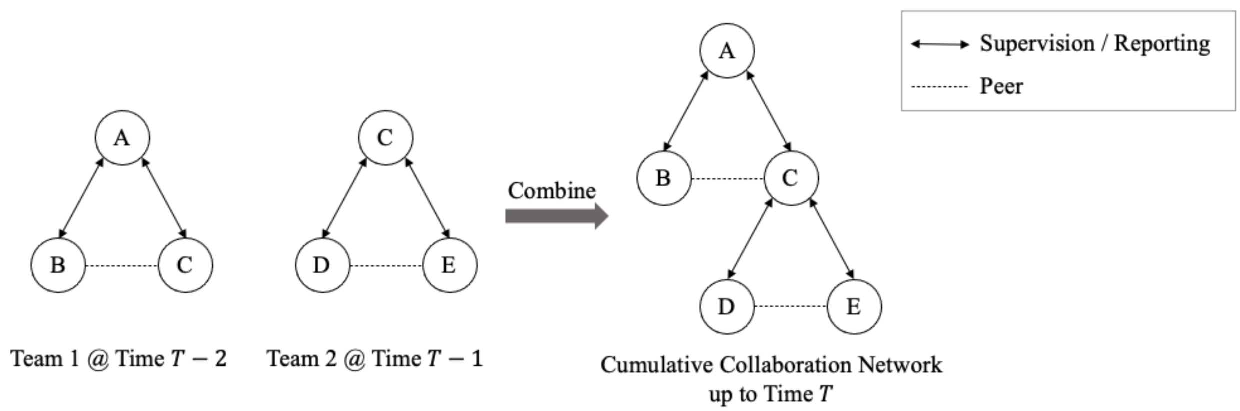

Figure 2: Generation of cumulative collaboration network

Then we extended the DeepWalk model [30] so that node similarities are learned with hierarchies in mind. The vanilla DeepWalk model uses a random walk approach and sequentially samples a node's neighborhood to generate the "contexts" of nodes. In the context of the collaboration network, the basic idea of DeepWalk is that individuals that have worked in the same team are more likely to have similar embeddings than those who have not.

Unlike DeepWalk, the proposed hierarchy2vec model adopts a hierarchically biased walk scheme to give different probabilities for the walk to traverse supervision, reporting, and peer ties in a hierarchical collaboration network. For example, if a team member tends to be influenced more by her supervisor than by her supervisees, then the walk should favor reporting ties than supervision ties so that the member's embedding is more similar to that of her supervisor. In the example cumulative collaboration network in Figure 2, consider that we are now at node C. The biased walk can decide the next node with the probability

p for supervision (Nodes D and E), q for reporting (Node A), and $1-p-q$ for peer (Node B) ties. Therefore, node C is likely to sample node A with the highest probability in case of $p \leq 1-p-q \leq q$.

After generating node sequences with the hierarchically biased walk, SkipGram model is then used to learn embeddings for all nodes (i.e., collaboration features) in the cumulative hierarchical collaboration network. For each node, its collaboration features are then concatenated with its individual feature to form its comprehensive set of node features.

3.4 Hierarchical node feature aggregation

After we generated node features, we aggregated the node features of individuals in the same team to generate the team-level representations. Specifically, we aggregated the node features in a hierarchical way, which gives the different importance to the nodes in each level of hierarchy.

Out of a cumulative hierarchical collaboration network, team is represented by a subgraph among its team members. Hierarchy2vec then learns the team's embedding by aggregating its team members' node features. To capture the hierarchical structure within a team, the hierarchy2vec model adopts a hierarchical way to aggregate node features in a bottom-up fashion (i.e., along the direction of reporting ties). Specifically, the aggregation starts from the lowest level of a team's hierarchy. Node features of team members at this level are first aggregated into one vector and then aggregated with their supervisor's node features. This process moves upward in the hierarchy until it reaches the top-most level. In the end, the aggregated features at the top-level become the team embedding, which integrates individual and collaboration features of all members in a team while giving higher weights to nodes with upper-level positions.

More formally, in a team T_t with a set of nodes N_t, assume that the team has L hierarchy levels. In the level $l (1 \leq l \leq L)$, the hidden state of a node $n \in N_t$ at the level l is denoted as $h_t^{l,n}$. This hidden state is the aggregated vector that combines the information collected up to the node n's apprentice nodes and the feature vector of the node itself. The apprentice nodes' information is aggregated by concatenating their hidden states if the current computation level is greater than or equal to 3. If the current level is at 2, the apprentice nodes are aggregated simply with the average of their feature vectors. Then, the aggregated apprentice nodes is again concatenated with the current node feature vector and assigned as a hidden state. This is formulated as follows:

$$h_t^{l,n} = \begin{cases} [h_t^{l-1,c_1} \| h_t^{l-1,c_2} \| \ldots \| h_t^{l-1,c_n} \| f_t^{l,n}] \cdot W^{(l)}, \{c_1, c_2, \ldots, c_n\} \in N_t^{(S)}(n) \text{ if } 3 \leq l \\ \\ [AVERAGE(f_t^{1,c}) \| f_t^{l,n}] \cdot W^{(2)}, c \in N_t^{(S)}(n) \text{ if } l = 2 \end{cases}$$

where $\|$ is the concatenation operator, $h_t^{l,n}$ is the hidden state of the node n in level l, $f_t^{l,n}$ is the feature vector of the node n, $W^{(l)}$ is the transformation that hierarchically aggregates the nodes at level l, and $N_t^{(S)}(n)$ is a set of $7n$'s neighbors connected by supervision ties. In the end, the hidden state of the topmost node becomes the team embedding z_t. Figure 3 illustrates the hierarchical aggregation of an example team with 8 nodes and 3 hierarchy levels.

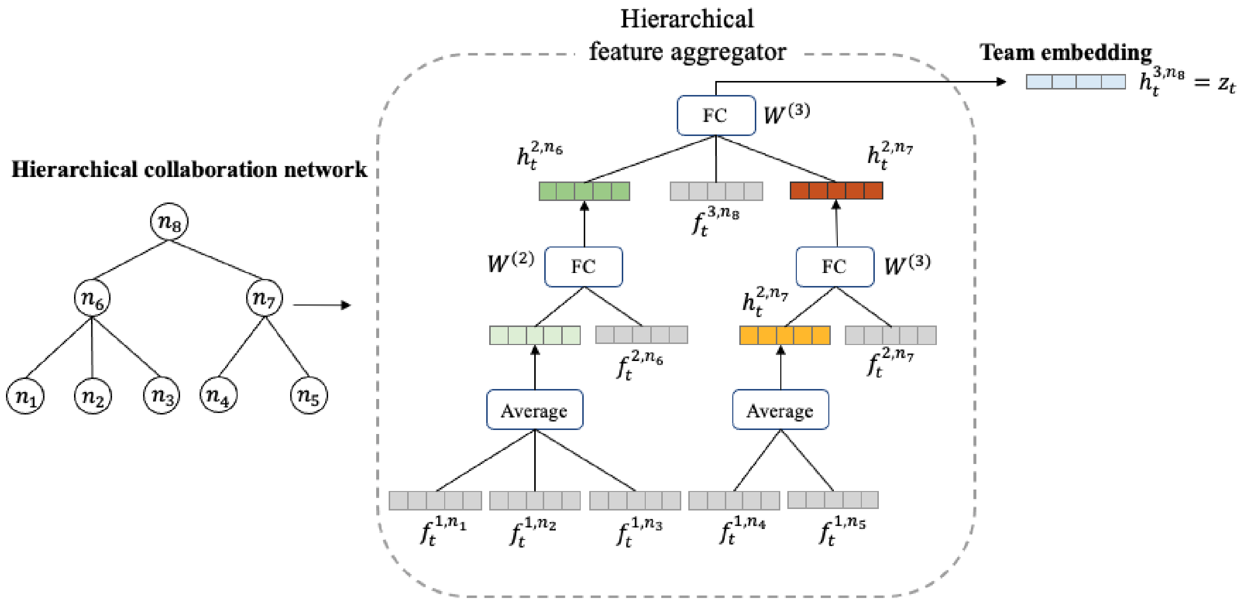

Figure 3: Hierarchical aggregation of node features for team embedding

4 Experiments

4.1 Data

Our experiments are based on a dataset of coaches and teams in the National Football League (NFL). We collected, cleaned and consolidated data from Pro Football Archives, Pro Football Reference, Pro Football History, and Wikipedia pages.

For each team in each season between 2002 and 2019, we first collected the team's performance data in the form of the number of wins and losses during regular seasons (a total of 16 games). We then collected data for coaches who took qualified coaching roles within each team during the 18 seasons. While there are many different types of roles within the coaching staff of an NFL team, we focused on head coaches, coordinators for offensive/defensive/special teams, and position coaches because these positions have clear responsibilities and exist in most teams. More importantly, the hierarchical structure among these coaches is very clear.

4.2 Network construction

A coaching team includes all coaches who worked for the same NFL team during a specific season. In the NFL coach hierarchy, head coaches are at the top of the structure with the greatest power and responsibility. At the second level of the hierarchy, defensive coordinators, offensive coordinators, and special teams coordinators work directly under the head coach. Each coordinator is responsible for one area of the team's play. Position coaches are located at the third level of the hierarchy. Each position coach is responsible for players at one specific position on the field and reports to one coordinator. For example, a quarterback coach is responsible for the play of quarterbacks on the field. Because quarterbacks are offensive players, a quarterback coach reports to the offensive coordinator, who is directly supervised by the head coach. Figure 4 illustrates the hierarchical structure of a typical NFL coach team based on coaches' titles.

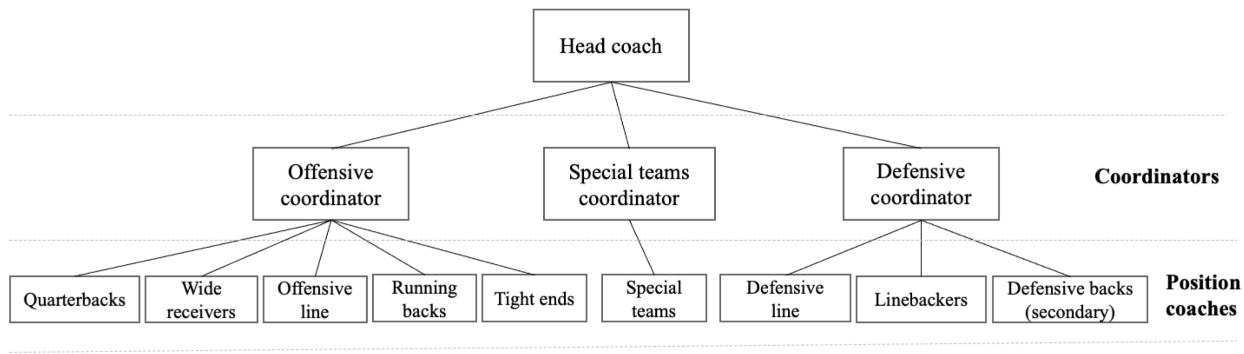

Figure 4: An illustration of NFL coach team structures.

Such three-level hierarchical structures allow us to construct a hierarchical collaboration network with hierarchical relationships among coaches who served the same team in the same season. In a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, $\mathcal{V}$ is a set of nodes representing coaches who served teams during 2002-2019 seasons and $\{\mathcal{E}^{(\mathcal{S})}, \mathcal{E}^{(\mathcal{R})}, \mathcal{E}^{(\mathcal{P})}\} \in \mathcal{E}$ is a set of edges representing the hierarchical collaborations among coaches connected by supervision, reporting, and peer ties. We represented each node as a tuple of serving year, serving team, and coach name $(y, t, c) \in \mathcal{V}$, which is a coach c who served a team t in year y. Each node has features represented as a vector $f_{y,t,c}$. The supervision edges $\mathcal{E}^{(\mathcal{S})}$ includes edges from a head coach to coordinators and from coordinators to their position coaches, the reporting edges $\mathcal{E}^{(\mathcal{R})}$ are the reciprocal direction of the supervision edges, and the peer edges $\mathcal{E}^{(\mathcal{P})}$ are connected among coordinators and among position coaches in the same group. Therefore, a connected component in the NFL coach collaboration network is a team $T_{y,t}$ consisted of coaches who worked at NFL team t in year y.

4.3 Experiment design

An NFL team's performance in a season can be measured in different ways. For example, a successful season can be defined as winning the Super Bowl, qualifying for play-offs, or records in the regular seasons. Given the complexity in defining a success, we experimented on predicting failures. We defined the team's season to be a failure if the team failed to win 50% of its regular season game or the head coach was fired in the middle of the season. To predict team failures in year y, we learned coach features using the data up to year $y - 1$. For training our models, we split the 18 years of NFL dataset into training (seasons between 2002 and 2015), validation (seasons in 2016 and 2017), and test (seasons in 2018 and 2019) sets. Table 1 shows the statistics of training, validation, and test collaboration networks.

We experimented on the following feature sets to illustrate the performance of hierarchy2vec.

- Feature set 1: Individual features
- Feature set 2: Individual features + Collaboration features (DeepWalk)

Table 1: NFL collaboration network statistics

	Training	**Validation**	**Test**
Number of coaching teams	448	64	64
Number of coaching teams with failures	196	27	33
Number of nodes	5,493	806	804
Number of edges	6,817	1,065	1,154

- Feature set 3: Individual features + Collaboration features (Hierarchical walk)

For each coach, individual features include the total years he coached in the NFL (as a measure of experience) and recent winning percentages in the NFL (as an indicator of his recent performance). Since we valued the overall and best performance, we used the average and the highest winning percentages during the recent 5 years of each coach's NFL coaching career.

Recall that collaboration features are node embeddings from the hierarchical collaboration network. When learning node embeddings using hierarchy2vec's hierarchically biased walk (Feature set 3), we set the different probabilities for traversing supervision, peer, and reporting edges with the probability ratio of 1:3:5. As a comparison, we also included embeddings learned with the random walk as in DeepWalk. For both types of the walk, we used the window size of 7, walk length of 10, 100 epochs, and embedding size of 30. The hyper-parameters are selected by searching for the best performing set near the values that are used in DeepWalk paper [30].

In order to evaluate the contribution of hierarch2vec's hierarchical aggregation of node features, we included a baseline non-hierarchical model, which does not consider the hierarchy structures in teams. Specifically, given a team T_t with N team members, a matrix of node features is $F_t \in \mathbb{R}^{N \times f}$, where f is the dimension of node features, the non-hierarchical aggregation generates a team embedding z_t of team T_t as follows:

$$\tilde{F}_t = AVERAGE(F_t)$$
$$z_t = \tilde{F}_t W^{(1)},$$

where $\tilde{F}_t \in \mathbb{R}^{1 \times f}$, $W^{(1)} \in \mathbb{R}^{f \times d}$, and $z_t \in \mathbb{R}^{1 \times d}$ with the pre-defined team embedding size d. In sum, the non-hierarchical approach first averages all node features; then, we used a fully connected neural network layer that transforms the averaged node features into a lower-dimensional vector space of size d.

While training the hierarchical aggregation model, we used $d = 16$ as the team embedding size. We used the early stopping strategy to stop training when the validation loss does not decrease for 50 epochs to prevent model overfitting. We optimized the model parameters using binary cross-entropy loss.

4.4 Results

We used Area Under the Curve (AUC) as the performance evaluation metric. Since exploring the optimized model parameters depends on the initial settings, we used 10 different random seeds for setting the randomized initial parameters for all models and reported the averaged AUCs over 10 repeated training.

Table 2 compares test AUCs of models with three feature sets as inputs and using non-hierarchical and hierarchical feature aggregations. The hierarchical aggregation model with Feature set 3, i.e., hierarchy2vec, achieves the highest AUC (Test AUC = 0.653). When comparing models in the same column, we can see that collaboration features generated by the proposed hierarchically biased walk outperforms DeepWalk's random walk. In addition, hierarchical aggregation works better than non-hierarchical aggregation, no matter which feature set is used. All these improvements are statistically significant with p-values below 0.05.

Table 2: Model comparisons

	Non-hierarchical aggregation	Hierarchical aggregation
Feature set 1	0.572	0.631
Feature set 2 (DeepWalk)	0.585	0.597
Feature set 3 (Hierarchical walk)	0.600	**0.653**

5 Summary

This paper proposed hierarchy2vec, a graph representation learning model designed for teams with hierarchical structures. The model consists of two major modules: a node embedding learning module that leverages hierarchically biased walk in hierarchical collaboration networks, and an end-to-end team embedding learning model that aggregates node features in a hierarchical way.

Through experiments on predicting NFL coaching teams' performance, we demonstrated the superior performance of hierarchy2vec, as well as the unique contribution of its two modules. The results suggest that beyond individual members' prior experience and performance, how a team is assembled also matters to predicting its performance. Moreover, when predicting a team's performance, one needs to consider the hierarchical structure within the team, and pay more attention to the value of those holding higher level positions. We hope our findings can have implications for managerial decisions on how to create high-performing teams.

This study also has limitations that we plan to address in future research. First, the experiment is limited to NFL because the three-level hierarchy in coaching teams is clearly defined. Using data from another domain would help the generalizability. Second, as a robustness check, we will also vary model parameters and experiment settings. Third, ties in hierarchical collaboration networks are unweighted. It would be interesting to see if a specific weighting scheme (e.g., by collaboration frequency or recency) can further improve the performance of hierarchy2vec. Last, but not least, the goal of this paper is to demonstrate the value of a team's hierarchical structures in predicting the team's performance. If the objective is to achieve the best predictions for NFL teams, a model needs to consider many other factors beyond coaching teams, such as the abilities of and the "chemistry" among players.

References

[1] Luigi Aldieri, Maxim Kotsemir, and Concetto Paolo Vinci. The impact of research collaboration on academic performance: An empirical analysis for some European countries. *Socio-Economic Planning Sciences*, 62:13–30, June 2018.

[2] Victor Amelkin, Omid Askarisichani, Young Ji Kim, Thomas W. Malone, and Ambuj K. Singh. Dynamics of collective performance in collaboration networks. *PLOS ONE*, 13(10):e0204547, October 2018.

[3] Soner Aydın and Soner Aydin. *Link prediction models for recommendation in academic collaboration network of Turkey*. Thesis, 2017.

[4] Albert Banal-Estañol, Mireia Jofre-Bonet, and Cornelia Lawson. The double-edged sword of industry collaboration: Evidence from engineering academics in the UK. *Research Policy*, 44(6):1160–1175, July 2015.

[5] Yaneer Bar-yam. *COMPLEX SYSTEMS AND SPORTS: Complex Systems insights to building effective teams*. 2003.

[6] Nesserine Benchettara, Rushed Kanawati, and Céline Rouveirol. A supervised machine learning link prediction approach for academic collaboration recommendation. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 253–256, New York, NY, USA, September 2010. Association for Computing Machinery.

[7] A G Biichner, W Dubitzky, A Schuster, P Lopes, P G O'Doneghue, J G Hughes, D A Bell, K Adamson, J A White, JoM C C Andersonv, and M D Mulvenna. Corporate Evidential Decision Making in Performance Prediction Domains. page 8, 1997.

[8] Vicente P. Guerrero Bote, Carlos Olmeda-Gómez, and Félix de Moya-Anegón. Quantifying the benefits of international scientific collaboration. *Journal of the American Society for Information Science and Technology*, 64(2):392–404, 2013.

[9] HongYun Cai, Vincent W. Zheng, and Kevin Chen-Chuan Chang. A Comprehensive Survey of Graph Embedding: Problems, Techniques, and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637, September 2018.

[10] Mei Cao and Qingyu Zhang. Supply chain collaboration: Impact on collaborative advantage and firm performance. *Journal of Operations Management*, 29(3):163–180, 2011.

[11] Amar Dhand, Douglas A. Luke, Bobbi J. Carothers, and Bradley A. Evanoff. Academic Cross-Pollination: The Role of Disciplinary Affiliation in Research Collaboration. *PLOS ONE*, 11(1):e0145916, January 2016.

[12] Andrew Fast and David Jensen. The NFL Coaching Network: Analysis of the Social Network among Professional Football Coaches. *AAAI Fall Symposium - Technical Report*, page 8.

[13] Lori Fewster-Thuente and Barbara Velsor-Friedrich. Interdisciplinary Collaboration for Healthcare Professionals. *Nursing Administration Quarterly*, 32(1):40–48, January 2008.

[14] Fahimeh Ghasemian, Kamran Zamanifar, Nasser Ghasem-Aqaee, and Noshir Contractor. Toward a better scientific collaboration success prediction model through the feature space expansion. *Scientometrics*, 108(2):777–801, August 2016.

[15] Linxia Gong, Xiaochuan Feng, Dezhi Ye, Hao Li, Runze Wu, Jianrong Tao, Changjie Fan, and Peng Cui. OptMatch: Optimized Matchmaking via Modeling the High-Order Interactions on the Arena. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, pages 2300–2310, New York, NY, USA, August 2020. Association for Computing Machinery.

[16] Aditya Grover and Jure Leskovec. node2vec: Scalable Feature Learning for Networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, New York, NY, USA, August 2016. Association for Computing Machinery.

[17] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 1025–1035, Red Hook, NY, USA, December 2017.

[18] Julia E. Hoch and Steve W. J. Kozlowski. Leading virtual teams: Hierarchical leadership, structural supports, and shared team leadership. *Journal of Applied Psychology*, 99(3):390–403, 2014.

[19] Kyujin Jung and Minsun Song. Linking emergency management networks to disaster resilience: bonding and bridging strategy in hierarchical or horizontal collaboration networks. *Quality & Quantity*, 49(4):1465–1483, July 2015.

[20] Elizabeth Keating and Chiho Sunakawa. Participation cues: Coordinating activity and collaboration in complex online gaming worlds. *Language in Society*, 39(3):331–356, 2010.

[21] Thomas N Kipf and Max Welling. Semi-supervised Classification with Graph Convolutional Networks. 2017.

[22] Jenny M. Lewis, Sandy Ross, and Thomas Holden. The how and why of academic collaboration: disciplinary differences and policy implications. *Higher Education*, 64(5):693–708, November 2012.

[23] Yongjun Li, Lizheng Wang, and Feng Li. A data-driven prediction approach for sports team performance and its application to National Basketball Association. *Omega*, 98:102123, January 2021.

[24] Chien Hsiang Liao. How to improve research quality? Examining the impacts of collaboration intensity and member diversity in collaboration networks. *Scientometrics*, 86(3):747–761, November 2010.

[25] Jukka Majava, Ville Isoherranen, and Pekka Kess. Business Collaboration Concepts and Implications for Companies. *International Journal of Synergy and Research*, 2(1-2), 2013.

[26] Bryan A McCullick, Jeremy Elliott, and Paul G Schempp. An analysis of National Football League coaching trees and the network they comprise. *International Journal of Sports Science & Coaching*, 11(1):4–15, February 2016.

[27] Ajay Mehra, Brett R. Smith, Andrea L. Dixon, and Bruce Robertson. Distributed leadership in teams: The network of leadership perceptions and team performance. *The Leadership Quarterly*, 17(3):232–245, June 2006.

[28] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. January 2013.

[29] Buraj Patrakosol and David L. Olson. How interfirm collaboration benefits IT innovation. *Information & Management*, 44(1):53–62, January 2007.

[30] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. DeepWalk: online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '14, pages 701–710, New York, NY, USA, August 2014. Association for Computing Machinery.

[31] Leonardo F.R. Ribeiro, Pedro H.P. Saverese, and Daniel R. Figueiredo. struc2vec: Learning Node Representations from Structural Identity. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 385–394, New York, NY, USA, August 2017. Association for Computing Machinery.

[32] Anna Sapienza, Palash Goyal, and Emilio Ferrara. Deep Neural Networks for Optimal Team Composition. *Frontiers in Big Data*, 2, 2019.

[33] Christian Tröster, Ajay Mehra, and Daan van Knippenberg. Structuring for team success: The interactive effects of network structure and cultural diversity on team potency and performance. *Organizational Behavior and Human Decision Processes*, 124(2):245–255, July 2014.

[34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is All You Need. 2017.

[35] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. February 2018.

[36] Shan Wang and Norm Archer. Business-to-business collaboration through electronic marketplaces: An exploratory study. *Journal of Purchasing and Supply Management*, 13(2):113–126, March 2007.

[37] Qi Wen, Maoshan Qiang, and Peter Gloor. Speeding up decision-making in project environment: The effects of decision makers' collaboration network dynamics. *International Journal of Project Management*, 36(5):819–831, July 2018.

[38] Linchuan Xu, Jiannong Cao, Xiaokai Wei, and Philip S. Yu. Network Embedding via Coupled Kernelized Multi-Dimensional Array Factorization. *IEEE Transactions on Knowledge and Data Engineering*, 32(12):2414–2425, December 2020.

[39] Daokun Zhang, Jie Yin, Xingquan Zhu, and Chengqi Zhang. Network Representation Learning: A Survey. *IEEE Transactions on Big Data*, 6(1):3–28, March 2020.

[40] Zhiya Zuo and Kang Zhao. Understanding and predicting future research impact at different career stages—A social network perspective. *Journal of the Association for Information Science and Technology*, 72(4):454–472, 2021. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.24415.