

# Machine Learning on Graphs: Towards the Development of an Intelligent Tutor in a Serious Game Aimed at Introducing Python

Matthieu Branthôme  
matthieu.branthome@irisa.fr

David Gross-Amblard  
david.gross-amblard@irisa.fr

September 20, 2024

## 1 Context

Since 2019, the teaching of computer programming has been introduced into secondary school curricula in France. In lower secondary school (grade 7 to 9), students start learning through block programming languages such as Scratch. From grade 10, the Python language is then introduced in mathematics and “Sciences Numériques et Technologie” (SNT) courses. The *Pirates* software (<https://pirates.fr>) was developed in 2021 [3] to offer a first approach to the Python language for grade 10 students, while easing the transition from block languages. This online application is a platformer game set in the world of pirates, in which students have to control a character using a Python script. The application, which to date was played over 200,000 times, is able to trace learners’ interactions with the environment (content displayed, progress through levels, code submitted, copy-paste, errors, etc.). These traces reveal that, in the classroom, students encounter difficulties as they progress, and regularly seek help from their teacher. In this context, the development of an intelligent tutor capable of delivering on demand tailored assistance to learners would be able to relieve teachers in the field. Automated feedback systems [4] generated from learners’ previous data have already been the subject of research in programming learning environments [8, 9, 7]. However, to the best of our knowledge, no one has explored the use of machine learning on graphs (e.g., GNN [10]) to address this problem.

## 2 Internship subject

During this internship, the student will :

1. Using the traces collected in the form of tabular data in a relational database, model the learners’ interactions with the application using Property Graph [1, 5] and store them in a suitable graph database [2] (e.g., Neo4j);
2. Train artificial intelligence models using machine learning techniques on graphs that are capable of link prediction [6]. The aim is to be able to predict the learner’s next action in the application at a given time;
3. To evaluate to what extent it would be possible to integrate this model into the *Pirates* application in order to develop an intelligent tutor capable of helping learners in an automated way by indicating the next action to be performed.

The trainee must be a computer science student at Master 2 level. He/she should have general skills in data science and machine learning. Knowledge of machine learning on graphs, in particular GNN, would be appreciated.

### 3 Environment

This internship will take place in the IRISA laboratory at the University of Rennes in the DRUID team. This team’s research focuses on data and knowledge management. The trainee will be based, at his/her convenience, in the DRUID team offices on the IRISA site in Rennes or Lannion.

### References

- [1] ANGLES, R. The property graph database model. In *AMW* (2018).
- [2] BESTA, M., GERSTENBERGER, R., PETER, E., FISCHER, M., PODSTAWSKI, M., BARTHEL, C., ALONSO, G., AND HOEFLER, T. Demystifying graph databases: Analysis and taxonomy of data organization, system designs, and graph queries. *ACM Computing Surveys* 56, 2 (2023), 1–40.
- [3] BRANTHÔME, M. Pyrates: Design and evaluation of a serious game aimed at introducing python programming and easing the transition from blocks. *ACM Trans. Comput. Educ.* 24, 1 (feb 2024).
- [4] DEEVA, G., BOGDANOVA, D., SERRAL, E., SNOECK, M., AND DE WEERDT, J. A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education* 162 (2021), 104094.
- [5] HOGAN, A., BLOMQUIST, E., COCHEZ, M., D’AMATO, C., MELO, G. D., GUTIERREZ, C., KIRrane, S., GAYO, J. E. L., NAVIGLI, R., NEUMAIER, S., ET AL. Knowledge graphs. *ACM Computing Surveys (Csur)* 54, 4 (2021), 1–37.
- [6] KUMAR, A., SINGH, S. S., SINGH, K., AND BISWAS, B. Link prediction techniques, applications, and performance: A survey. *Physica A: Statistical Mechanics and its Applications* 553 (2020), 124289.
- [7] MARWAN, S., AKRAM, B., BARNES, T., AND PRICE, T. W. Adaptive immediate feedback for block-based programming: Design and evaluation. *IEEE Transactions on Learning Technologies* 15, 3 (2022), 406–420.
- [8] PRICE, T. W., DONG, Y., AND LIPOVAC, D. Isnap: Towards intelligent tutoring in novice programming environments. In *SIGCSE Technical Symposium on Computer Science Education* (2017), ACM, p. 483–488.
- [9] RIVERS, K., AND KOEDINGER, K. R. Data-driven hint generation in vast solution spaces: a self-improving python programming tutor. *International Journal of Artificial Intelligence in Education* 27, 1 (2017), 37–64.
- [10] WU, Z., PAN, S., CHEN, F., LONG, G., ZHANG, C., AND YU, P. S. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems* 32, 1 (2021), 4–24.