

# Utilizando aprendizado por representação para a classificação de laços sociais da IoT

Jamisson J. Junior  
Thiago S. Figueiredo  
jamisson.junior@aluno.ufop.edu.br  
thiago.figueiredo@aluno.ufop.edu.br  
Universidade Federal de Ouro Preto  
João Monlevade, MG

Ramon Lopes  
rlopes@ufrb.edu.br  
Universidade Federal do Recôncavo  
da Bahia  
Cruz das Almas, BA

Bruno P. Santos  
Luiz C. B. Torres  
bruno.ps@ufba.br  
luiz.torres@ufop.edu.br  
Universidade Federal de Ouro Preto  
Universidade Federal da Bahia  
Salvador, BA

## Abstract

A Internet of Things (IoT) tem sido marcada pelas interações entre dispositivos que cooperam para realizar atividades. Este cenário viabiliza o paradigma o Social IoT (SIoT), onde múltiplos tipos de relacionamentos e confiabilidade podem ser estabelecidos entre dispositivos. Neste artigo, abordamos as questões de como modelar laços sociais em IoT e na proposição de modelos para, automaticamente, classificar e prever relações em SIoT. Assim, propomos a utilização de aprendizado por representação para classificar diferentes tipos de laços sociais em SIoT. Para isso, utiliza-se como estratégias para classificação Graph Neural Networks (GNN) ou Algoritmos Tradicionais de Classificação (ATC). Como resultados, mostramos o compromisso no uso de GNN e ATC em diferentes cenários.

**Keywords:** IoT, SIoT, Aprendizado por representação, Machine Learning

## 1 Introdução

A Internet das Coisas (do inglês *Internet of Things (IoT)*) engloba diversas tecnologias e dispositivos que já estão presentes em nosso cotidiano, os quais usam a estrutura de endereçamento da Internet e protocolos de comunicação. Essas características possibilitam que objetos interajam e colaborem por objetivos em comum. Com isso, surge o paradigma Internet das Coisas Sociais (do inglês *Social IoT (SIoT)*) [1] que aplica princípios das redes sociais à IoT, justamente para explorar o potencial das relações entre dispositivos. Com isso, a estrutura da SIoT pode ser moldada para se obter navegabilidade na rede tal como feito de modo escalável em redes sociais humanas, o que é importante para IoT na descoberta de dispositivos, serviços e recursos de maneira distribuída.

Neste sentido, existem ao menos dois desafios na SIoT que podem ser explorados: a identificação dos diferentes tipos de laços sociais que podem existir e a sua classificação. A

complexidade desses desafios está na falta de consenso na definição dos tipos de laços sociais na SIoT e na dinamicidade em que nós (dispositivos IoT) e laços (relacionamentos) aparecem e desaparecem ao longo do tempo. Recentes avanços na área de aprendizado de representação (do inglês *representation learning*) tem possibilitado o processamento de dados representados no domínio dos grafos através de técnicas como *node embeddings* e *Graph Neural Networks (GNN)*. Entretanto, no melhor de nossos conhecimentos o uso de *node embeddings* e GNNs para classificação das relações da SIoT ainda é pouco explorado.

Neste artigo, propomos a utilização de aprendizado por representação para classificação dos tipos de relações sociais da IoT. O algoritmo criado recebe como entrada um grafo representando a rede social IoT. Em seguida, a entrada é processada utilizando GNN ou *node embeddings* combinado com Algoritmos Tradicionais de Classificação (ATC). Como saída, é produzido um modelo para classificação dos enlaces sociais ainda não conhecidos. Em nossos experimentos, usamos como base de dados<sup>1</sup> a interação entre 16.216 dispositivos da IoT dispostos na cidade de Santander, localizada na Espanha, e comparamos com quatro diferentes técnicas de aprendizado de máquina. Como resultado, GNN se mostrou mais rápido na etapa de treinamento e classificação, enquanto apresenta métricas  $F1-\{macro, micro\}$  0.61 e 0.88, respectivamente. Ao usar ATC, o treinamento e classificação são mais lentos que GNN, ao passo que as métricas  $F1-\{macro, micro\}$  alcançam valores entre 0.92 e 0.95, respectivamente.

O restante do artigo está organizado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados. Na Seção 3, definimos o problema e descrevemos as abordagens de classificação utilizadas. Os ambiente de experimentação e a base de dados são apresentados na Seção 4. Os resultados são discutidos na Seção 5. Por fim, Seção 6 apresenta a conclusão.

## 2 Trabalhos relacionados

### 2.1 Internet das Coisas Sociais e os tipos de relações entre dispositivos

SIoT foi formalizada por [1], em que explorou-se a capacidade dos dispositivos da IoT em formar redes sociais de

In: II Concurso de Trabalhos de Iniciação Científica (CTIC 2022), Curitiba, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2022.  
© 2022 SBC – Sociedade Brasileira de Computação.  
ISSN 2596-1683

<sup>1</sup>Disponível em: <http://www.social-iot.org/index.php?p=downloads>.

forma autônoma baseada em regras dos seus proprietários. Após isso, pesquisas se concentraram em identificar políticas, métodos e técnicas para o estabelecimento de relação entre os dispositivos sem a necessidade de intervenção humana.

**Table 1.** Principais relações sociais da SIoT [8]

Relação IoT	Tipo de relação entre dispositivos:
<i>Parental object relationship (POR)</i>	Do mesmo tipo, modelo e marca;
<i>Ownership object relationship (OOR)</i>	Pertencentes ao mesmo proprietário;
<i>Co-location object relationship (C-LOR)</i>	Frequentemente no mesmo local; e se encontram de forma intermitente,
<i>Social object relationship (SOR)</i>	Em que seus donos entram em contato durante a rotina diária;
$SOR_2$	Variante da SOR que estabelece conexões entre dispositivos público móveis e privado móveis.

Nesse sentido, [8] propuseram 4 tipos de relacionamentos sociais que os dispositivos poderiam realizar uns com os outros, vide Tabela 1. Eles levaram em consideração não apenas que os dispositivos podem ser móveis (e.g. carros) ou estáticos (e.g. impressora), como também se pertencem a um domínio público ou são bens privados de um usuário. No melhor de nossos conhecimentos, a utilização de modelos de aprendizado de máquina capazes de classificar os relacionamentos entre os dispositivos da SIoT ainda é uma necessidade em aberto dentro literatura.

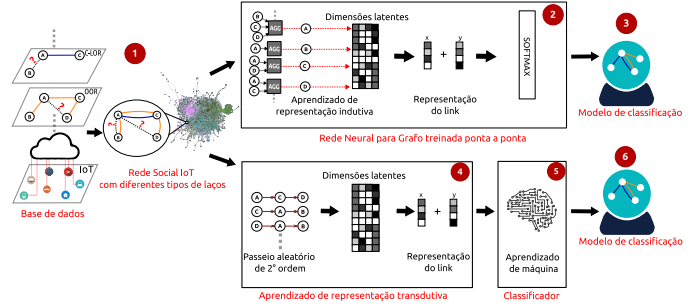
## 2.2 Aprendizado de representação para grafos

Recentemente diversas pesquisas vem sendo realizadas para entender como modelos de aprendizado de máquina podem auxiliar no estudo de redes sociais. Paradigmas focados em GNN têm mostrado um alto potencial em predição e classificação de conexões [7], de forma que parte do aprendizado está em identificar representações de baixa dimensão para os nós da rede. Nessa etapa, duas abordagens podem ser consideradas: o aprendizado transdutivo e o indutivo.

O aprendizado transdutivo consiste em aprender a incorporação de cada nó separadamente, aproveitando padrões baseados em fatoração de matrizes [4] como, por exemplo, o node2vec [3]. A limitação desses métodos está em classificar nós não vistos na etapa de treino e sua complexidade cresce com o tamanho da rede [10]. Por sua vez, o aprendizado indutivo utiliza funções de agregação capazes de reconhecer parâmetros globais do grafo [2, 9]. Dessa forma, modelos indutivos como o GraphSage [5] são capazes de reconhecer a inserção de novos nós. O tópico seguinte detalhará como essas abordagens foram utilizadas nesse trabalho.

## 3 Fluxo de execução

Para o problema em questão, seja  $C$  o conjunto de classes existentes para os relacionamentos entre nós em uma rede SIoT. Neste trabalho, um rede SIoT é representada como um grafo  $G = (V, E, r)$  com um conjunto  $V$  de vértices para representar os nós, um conjunto  $E \subseteq V \times V$  de arestas para



**Figure 1.** Fluxo de execução. Etapas (1-3) usam GNN e Etapas (1, 4-6) utilizam ATCs para realizar a classificação.

representar as interações entre os nós da rede e uma função  $r : E \rightarrow C$  para mapear uma aresta para uma classe de relacionamento. O problema consiste em, dada uma aresta  $e \in E$ , prever a classe  $c$  da aresta, em outras palavras, precisamos aprender uma função de aproximação da função  $r$ .

Neste sentido, este trabalho busca utilizar técnicas de aprendizado de máquina para encontrar a função de aproximação  $r$  utilizando o aprendizado de representação da aresta  $e$  como entrada para algoritmos de aprendizado de máquina. O fluxo de execução dos algoritmos é ilustrada na Figura 1. Os estágios compreendidos entre 1 e 3 mostram o passo a passo utilizando GNN para gerar o modelo de classificação. Por sua vez, as etapas 1, 4, 5 e 6 ilustram um processamento alternativo, o qual permite que Algoritmos Tradicionais de Classificação sejam utilizados para produzir o modelo de classificação. O processo se inicia na etapa 1, em que as diferentes redes de relacionamentos sociais dos dispositivos são integradas em um único grafo SIoT, onde: os nós representam os dispositivos e as arestas representam os múltiplos relacionamentos entre os pares de dispositivos.

Após a geração do grafo SIoT na etapa 1, a etapa 2 emprega um modelo GNN para realizar o aprendizado ponta-a-ponta, isto é, tanto a etapa de aprendizado de representação quanto a classificação são executadas conjuntamente. Por sua vez, a segunda abordagem emprega o aprendizado de representação na etapa 4 e classificação utilizando ATCs na etapa 5.

## 4 Experimentos

Nossos experimentos foram projetados para responder as seguintes questões:

- Q1. É possível classificar os tipos de laços sociais em SIoT?
- Q2. Como a escolha da abordagem afeta o desempenho?

Para responder a essas questões, utilizamos uma base de dados disponível na literatura e vários modelos de classificação em nossos experimentos conforme detalhado a seguir.

### 4.1 Base de dados

A base de dados utilizada modela a interação entre 16.216 dispositivos IoT reais e sintéticos dispostos na cidade de Santander, localizada na Espanha (vide Tabela 1). Processamos a

base de dados para construir uma rede de relacionamentos para cada tipo de conexão IoT, cujos rótulos são OOR, C-LOR, SOR e SOR<sub>2</sub>; esses rótulos compõem o conjunto  $C$  com base na definição do problema. Por simplicidade, para cada combinação possível entre os  $|C|$  rótulos existentes, criamos um novo rótulo. Por exemplo, se, entre um par de vértices, houver apenas duas arestas com rótulos OOR e C-LOR, um novo rótulo é criado para representar esses dois relacionamentos no grafo simples, por exemplo OOR+C-LOR. Isso possibilita simplificar o treinamento do modelo, onde é possível definir classes que são independentes entre si.

A base de dados foi dividida aleatoriamente em 70% para treino, 20% validação e 10% teste. Essa divisão foi realizada de forma a manter aproximadamente a mesma frequência das classes nos três conjuntos, isto é, empregamos um particionamento estratificado. Desse modo, seis classes, duas delas resultantes da combinação de dois tipos de relacionamento, estão presentes na base de dados com sua porcentagem indicada entre parênteses: OOR (50%); C-LOR (20%); SOR<sub>2</sub> (18%); SOR (7%); C-LOR+OOR (4.5%); C-LOR+SOR (0.5%).

## 4.2 Metodologia de avaliação

Para a avaliação dos resultados utilizamos a métrica  $F1$  score, definida como a média harmônica entre as métricas precisão (*precision*) e revocação (*recall*). Uma vez que a base de dados apresenta um desbalanceamento entre classes, reportamos as versões *micro* e *macro* desta métrica. Enquanto a versão *macro* calcula a métrica F1 para cada classe separadamente e, então, calcula a média aritmética entre esses valores, a versão *micro*, por sua vez, calcula a métrica de forma global.

### 4.2.1 Classificadores e calibração de hiperparâmetros.

Em nossos experimentos, para evitar o vazamento de informações (do inglês *data leakage*), os hiperparâmetros foram calibrados com base no resultado da métrica  $F1$ -*micro* obtido utilizando o conjunto de validação. Por fim, reportamos os resultados do conjunto teste com base na melhor configuração de hiperparâmetros para cada classificador. A seguir, discutimos os parâmetros setados no *grid search* para a calibração dos hiperparâmetros de cada um dos classificadores.

**Classificador GNN.** Para sua implementação, utilizamos o modelo GraphSAGE da biblioteca DGL<sup>2</sup>, em que escolhemos a função média como função agregadora e, a função ReLU como função de ativação. A arquitetura da rede neural possui uma camada de entrada com  $d$  dimensões,  $h$  camadas ocultas de dimensão  $2d$  exceto a última camada oculta que possui dimensão  $d$  e, por fim, uma camada *softmax* de saída. Para fins de treinamento, utilizamos o algoritmo de otimização Adam. Para calibração de hiperparâmetros, nós investigamos valores para  $d \in \{8, 16, 32, 64\}$ ,  $h \in \{0, 1, 2\}$ , fator de decaimento  $wd \in \{1e-5, 1e-4, 1e-3, 1e-2\}$  e taxa de aprendizado  $\alpha \in \{1e-5, 1e-4, 1e-3, 1e-2\}$ .

**Classificadores ATC.** Utilizamos o *framework* node2vec<sup>3</sup> para o aprendizado de representação dos vértices, e os modelos Logistic Regression (LR), Random Forest (RF) e Support Vector Machine (SVM) como classificador. O *framework* node2vec possui os hiperparâmetros  $p$  e  $q$  que controlam a largura e a profundidade da caminhada aleatória respectivamente e  $d$  dimensões de representação. Nesse sentido, investigamos valores para  $p, q \in \{0.5, 1, 2\}$  e  $d \in \{8, 16, 32, 64\}$ , já os parâmetros *walk length* e *number of walks* foram fixados em 80 e 10, respectivamente conforme sugerido em [6].

## 5 Resultados

Nesta seção, respondemos cada um dos questionamentos levantados na Seção 4. O desempenho dos melhores modelos sobre a métrica  $F1$ -score é apresentado na Figuras 2.

### 5.1 Classificação dos tipos de laços sociais em SIoT

Para responder a questão Q1, comparamos o desempenho dos classificadores apresentados na Seção 4.2.1 com aquele obtido por um simples classificador aleatório denominado Randômico (RAN). A saber, RAN classifica uma aresta  $e$  como classe  $c$  de forma aleatória com probabilidade  $p_c$ . Em nossos experimentos,  $p_c$  é definido pela frequência de arestas da classe  $c$  no conjunto de treino.

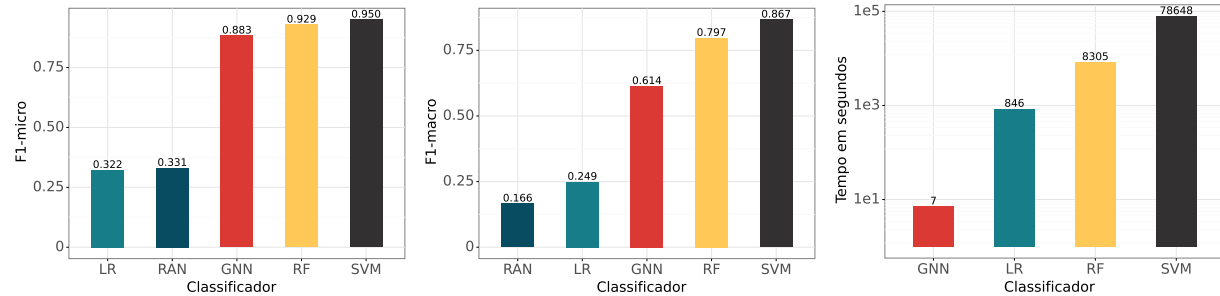
As Figuras 2a e 2b apresentam as métricas  $F1$ -*micro* e  $F1$ -*macro*, respectivamente, para cada classificador. É possível identificar que todos os classificadores superaram o classificador RAN nas métricas consideradas, com exceção do LR na métrica  $F1$ -*micro*. Dessa forma, lembrando a pergunta Q1, os resultados nos permitem concluir que é possível classificar os tipos de laços sociais em redes SIoT.

### 5.2 Efetividade na classificação

Para responder a questão Q2, avaliamos o desempenho usando GNN ou Algoritmos Tradicionais de Classificação (LR, RF e SVM). Com base nas Figuras 2a e 2b, podemos concluir que ao usar GNN melhores resultados são obtidos se comparado com LR em todas as métricas. Por outro lado, a abordagem GNN não é capaz de fornecer resultados superiores àqueles do RF e do SVM, mantendo um desempenho global de 92% do melhor modelo clássico. Não obstante, o SVM apresenta resultados superiores àqueles obtidos por RF. Apesar de não apresentar os melhores resultados na tarefa de classificação, o GNN despende um tempo para treinamento e classificação consideravelmente inferior ao seus pares, vide Figura 2c. Podemos concluir que GNN despendeu o menor tempo, aproximadamente 0.8% do tempo gasto por uma abordagem que utiliza modelos simples, como o LR.

<sup>2</sup><https://www.dgl.ai/>

<sup>3</sup><http://snap.stanford.edu/node2vec/>



(a) Desempenho sobre a métrica F1-micro (b) Desempenho sobre a métrica F1-macro (c) Tempo de treinamento e classificação

Figure 2

Uma análise mais detalhada dos resultados nos permite identificar que a etapa utilizando GNN não é capaz de classificar corretamente instâncias das três classes menos representativas na base de dados (i.e. SOR, C-LOR+OOR, C-LOR+SOR), alcançando, então, taxa de acerto próximo a zero. Desse modo, faz-se necessária a investigação de estratégias para lidar com o desbalanceamento de classes, seja do ponto de vista da base de dados seja do ponto de vista arquitetural de redes profundas. Além disso, os resultados superiores do SVM possui relação com o tamanho das dimensões escolhidas para a representação dos dispositivos, uma vez que esse modelo é conhecido por sua efetividade em altas dimensões.

Em resumo, os resultados dentro dos parâmetros testados, indicam que modelos como LR são pouco eficientes para identificar os padrões presentes dentro da rede de relacionamento da SIoT quando comparados com os outros classificadores utilizados. Além disso, o RF obteve o melhor desempenho dentre todos os modelos nas baixas dimensões, porém, o aumento da dimensão dos *embeddings* deteriora os resultados desse modelo. Por outro lado, o SVM obteve a maior acurácia sobretudo nas altas dimensões, no entanto, para isso necessitou de um tempo, no mínimo, 10 vezes maior de treinamento e classificação se comparado com seus pares. Por fim, embora o GNN não tenha atingido os melhores resultados, a vantagem desse modelo está no baixo tempo de treinamento e classificação, além da capacidade indutiva de reconhecer novos dispositivos adicionados na rede, sendo essa uma vantagem importante em um cenário dinâmico.

## 6 Conclusões

Os avanços na área da IoT possibilitam que dispositivos sejam capazes de realizar laços sociais uns com os outros, gerando então o paradigma SIoT. Por natureza, essas conexões podem representar múltiplos tipos de relacionamentos e obedecem às regras estabelecidas pelo proprietário. Entender como esses relacionamentos são estruturados dentro da rede, como também encontrar modelos capazes de classificar essas relações, é um problema em aberto dentro da literatura.

Nesse sentido, utilizar os avanços na área de aprendizado de máquina pode ser um caminho para encontrar técnicas

escaláveis e com capacidade de identificar, compreender e prever padrões sociais dos dispositivos. Por isso, nesse artigo utilizamos abordagens de GNN e ATC com aprendizado por representação para classificar os tipos de laços sociais da IoT. Para mostrar a efetividade dessas abordagens, comparamos os modelos na tarefa de classificar conexões de múltiplos tipos em uma rede SIoT baseada em dispositivos reais. Com os resultados pôde-se constatar que é possível classificar os tipos de laços sociais em redes SIoT utilizando modelos de aprendizado de máquina.

## References

- [1] Luigi Atzori, Antonio Iera, Giacomo Morabito, and Michele Nitti. 2012. The Social Internet of Things (SIoT) – When social networks meet the Internet of Things: Concept, architecture and network characterization. *Computer Networks* 56, 16 (2012), 3594–3608.
- [2] Gabriele Corso, Luca Cavalleri, Dominique Beaini, Pietro Liò, and Petar Veličković. 2020. Principal neighbourhood aggregation for graph nets. *arXiv preprint arXiv:2004.05718* 33 (2020), 13260–13271.
- [3] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *22nd ACM SIGKDD*. 855–864.
- [4] Chitrang Gupta, Yash Jain, Abir De, and Soumen Chakrabarti. 2021. Integrating Transductive and Inductive Embeddings Improves Link Prediction Accuracy. In *Proceedings of the 30th ACM International Conference on Information Knowledge Management (Virtual Event, Queensland, Australia) (CIKM '21)*. Association for Computing Machinery, New York, NY, USA, 3043–3047. <https://doi.org/10.1145/3459637.3482125>
- [5] William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. *ArXiv e-prints*.
- [6] Liang Hu, Gang Wu, Yongheng Xing, and Feng Wang. 2019. Things2Vec: Semantic modeling in the Internet of Things with graph representation learning. *IEEE Internet of Things Journal* (2019).
- [7] Abdullah Khanfor, Amal Nammouchi, Hakim Ghazzai, Ye Yang, Mohammad R Haider, and Yehia Massoud. 2020. Graph neural networks-based clustering for social internet of things. In *63rd MWSCAS*. IEEE.
- [8] Claudio Marche and Luigi Atzori. 2018. A Dataset for Performance Analysis of the Social Internet of Things. In *IEEE 29th PIMRC*.
- [9] Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. 2019. Weisfeiler and leman go neural: Higher-order graph neural networks. In *AAAI*.
- [10] Quentin Oliveau and Hichem Sahbi. 2018. From Transductive to Inductive Semi-Supervised Attributes for Ship Category Recognition. In *IEEE IGARSS*.