

A Link Prediction-Based Method Towards Lead Management

G. P. Brugalli¹, A. L. Gonçalves¹, A. S. Bordin¹, L. S. Artese¹

¹ Universidade Federal de Santa Catarina, Brazil
giuliobrugalli@gmail.com, a.l.goncalves@ufsc.br, andrea.bordin@ufsc.br,
artese.leticia@gmail.com

Abstract. Lead management is an essential part of the customer acquisition and retention stages. However, as the number of leads increases, data-driven management automation is critical in this process. In this context, the present work proposes a method that supports lead management to identify and recommend to the sales team, future interests of leads that already exist in an organization's database in order to acquire or retain customers. To fulfill this objective, the network representation learning and link prediction models are explored. A case study is presented to demonstrate the effectiveness of the proposed method. All generated models reached a value between 0.873 and 0.998 of ROC-AUC. However, the prediction models showed low accuracies, far from 1, the ideal value. Nevertheless, the method shows promise to be investigated in practice. For future work, a deep understanding of technical capabilities of network learning is suggested to obtain better results from link prediction models.

CCS Concepts: • **Computing methodologies** → **Machine learning algorithms**.

Keywords: lead management, network representation learning, link prediction, machine learning

1. INTRODUÇÃO

A análise de dados tem se tornado crucial para estratégias de aquisição e retenção de clientes. O acompanhamento e gestão dos *leads* é uma parte essencial para tais estratégias e, neste sentido, existe um investimento substancial na geração de *leads*. Todavia, estatísticas mostram que a maioria dos *leads* são ignorados e nunca contatados [Ohiomah et al. 2016].

O conceito de *leads* se define como um registro de interesse, de alguma pessoa, por algum produto ou serviço de uma organização. Esse registro normalmente possui informações básicas para que o potencial cliente (*lead*) seja submetido à estratégias de vendas [Ohiomah et al. 2016]. Uma pessoa pode ser um cliente e um *lead* ao mesmo tempo, pois além de já ter consumido algum produto ou serviço, ainda pode registrar interesse em continuar consumindo.

Entretanto, à medida que o número de *leads* aumenta, o processo de aquisição/retenção de clientes se torna mais complexo e pouco eficiente [Yu and Cai 2007; D'Haen et al. 2013]. Em um cenário com muitos *leads*, para que a seleção dos melhores seja pautada de forma inteligente e racional, otimizando a aquisição e retenção de clientes, D'Haen et al. [2016] afirmam que um sistema automatizado de gestão de *leads* é imprescindível. Segundo Gebert et al. [2002], a gestão de *leads* representa a consolidação, qualificação e priorização de contatos em potenciais clientes, provendo à equipe de vendas uma lista de *leads* qualificada e priorizada.

Sendo assim, este trabalho tem sua relevância ao apontar um possível método de gestão de *leads* a partir do conceito de predição de *links*. Visa ainda contribuir na identificação e recomendação, para o time de vendas, de futuros interesses de *leads* já existentes na base de dados de uma organização. Ademais, espera-se que os resultados obtidos possam ajudar na promoção de novas pesquisas em temas

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001

Copyright©2022 Permission to copy without fee all or part of the material printed in KDMiLe is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

correlatos.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Gestão de *leads*

A Gestão de *Leads* é responsável por consolidar, qualificar e priorizar os *leads* para identificar potenciais clientes, assim como apoiar equipes de vendas, principalmente quando o volume de *leads* cresce e a qualificação se torna complexa [Ohiomah et al. 2016].

Não existe de fato uma definição universal aceita sobre como qualificar um *lead*, pois as características levadas em consideração dependem da necessidade e do modelo de negócio [Monat 2011; Giacosa et al. 2022]. Estudos mostram que em empresas baseadas no modelo de *business-to-business*, representantes de vendas não entram em contato com cerca de 70% dos *leads* gerados pelo *marketing* [Michiels 2009]. Esse problema é conhecido como “buraco negro de *leads* de vendas”, representando a enorme quantidade de *leads* que ficam no limbo, sem serem contatados [Sabnis et al. 2013].

Como solução, a literatura mostra que o uso de tecnologia na área de *marketing* e vendas pode aumentar significativamente a eficiência da gestão de *leads* [Järvinen and Taiminen 2016]. Os resultados sugerem que isto tende a ocorrer quando um sistema de gestão de *leads* é adequadamente implementado por uma infraestrutura de Tecnologia da Informação e Comunicação (TIC) robusta e confiável [Giacosa et al. 2022].

2.2 Predição de *links*

Sistemas complexos têm atraído grande interesse de pesquisa nas últimas duas décadas. Assim, no século XXI nasce um novo campo de estudo: a Ciência das Redes, que modela esses sistemas como redes complexas [Barabasi and Posfai 2016; Molontay and Nagy 2019].

O formalismo matemático da Ciência das Redes é baseado na teoria de grafo, um subcampo da matemática que surgiu no século XVIII [Barabasi and Posfai 2016]. Segundo Vespignani and Caldarelli [2007], um grafo é um modo de codificar uma relação entre elementos de um sistema traçando vértices (elementos de um sistema) como pontos e as arestas (relações entre os elementos) como linhas entre estes.

A partir de grafos diversas tarefas podem ser executadas. Entre elas, encontra-se a Predição de *Links*, entendida como uma tarefa aplicada no contexto de redes complexas. Em uma definição formal, Martínez et al. [2016], explicam que dado um instante t no tempo, de uma rede não direcionada (em que as arestas não possuem direção), o problema de predição de *links* deduz um subconjunto de conexões ausentes no instante atual e que possivelmente serão formadas no tempo $t + \Delta$, sendo Δ (delta) uma variação da unidade de tempo.

Existem diversos algoritmos que se propõem a auxiliar na execução da tarefa de predição de *links*. Nos últimos anos, algoritmos baseados em aprendizado de incorporação (do inglês, *learning embeddings* - *LE*) têm ganhado destaque pela sua eficiência. Nestes algoritmos, o objetivo é mapear cada vértice para um vetor de baixa dimensão, de modo que esses vetores representem um resumo da estrutura da rede [Malek et al. 2021].

2.3 Redes Neurais Artificiais

Uma Rede Neural Artificial (do inglês *Artificial Neural Network* - ANN) é um sistema computacional não linear inspirado na estrutura, comportamento e habilidades de aprendizado de um cérebro biológico humano, buscando simular seu processamento de informações [Dharwal and Kaur 2016; Wu and Feng 2018].

No contexto da predição de *links*, redes neurais podem ser utilizadas para modelar possíveis soluções de aprendizado. Para tal, os algoritmos *Node2Vec* e *Metapath2Vec* são utilizados neste trabalho. De maneira geral, estes algoritmos se utilizam de ANNs para transformar as informações extraídas de um grafo em vetores que alimentam modelos de aprendizado para predição de *links*.

2.3.1 *Node2Vec*. O algoritmo *Node2Vec* aprende as características topológicas de cada vértice a partir de um processo conhecido como *node embedding* (NE), que mapeia cada vértice do grafo para um espaço vetorial, distribuído e contínuo, de baixa dimensionalidade. Cada vértice é transformado em um vetor representativo, ou seja, um vetor que contém as características daquele vértice. As relações geométricas no espaço vetorial refletem as propriedades e estrutura do grafo original. Os vetores são então usados como entrada para algoritmos de aprendizado de máquina (do inglês *Machine Learning* - ML) para extrair informações úteis para a construção de classificadores ou preditores. É importante destacar que o *Node2Vec* não distingue grafos homogêneos e heterogêneos [Peng et al. 2020; Amara et al. 2021].

Para realizar a tarefa de predição de *links* é preciso gerar vetores que correspondam às conexões, ou seja, vetores das arestas. Neste sentido, utiliza-se um operador binário entre vetores de dois vértices, resultando no vetor da aresta correspondente. Grover and Leskovec [2016] consideram quatro possíveis escolhas de operadores binários. Neste trabalho foi escolhido o operador *Weighted-L1* definido como $|u - v|$, sendo u e v representações vetoriais de vértices pertencentes ao grafo. O resultado é um vetor representativo da aresta entre u e v .

2.3.2 *Metapath2Vec*. O algoritmo *Metapath2Vec*, similar ao *Node2Vec*, também realiza o processo de NE utilizando o modelo de ANN *Skip-gram* para mapear um vetor representativo para cada vértice de um grafo em um espaço vetorial. Além de preservar a estrutura topológica, sua principal distinção está na semântica de um grafo com mais de um tipo de vértice, ou seja, um grafo heterogêneo [Dong et al. 2017].

No *Metapath2Vec* é possível utilizar os mesmos operadores binários do *Node2Vec* entre vetores dos vértices para gerar os vetores das arestas, produzindo assim, um modelo de aprendizado para realizar a tarefa de predição de *links*.

3. MÉTODO PROPOSTO

Este trabalho apresenta um método de análise de gestão de *leads* voltado à previsão de interesse de *leads* existentes por novas edições de produtos de uma determinada organização. O intuito é apoiar a otimização de ativação de *leads* pela equipe de vendas. Para alcançar este objetivo, utilizam-se técnicas de aprendizado de representação de espaços vetoriais de redes e predição de *links* no contexto da gestão de *leads*. A Figura 1 fornece uma visão geral do método e de suas 5 etapas. A seguir, cada etapa é descrita em mais detalhes.

3.1 Etapa 1: Pesquisa e coleta de dados

Nesta primeira etapa é realizada uma pesquisa em todas as bases de dados de uma determinada organização sobre o histórico referente aos *leads* (como o *e-mail*, valor gasto com a organização e edição de produtos com registro de interesse) e informações dos produtos (como categoria do produto, preço, número de edições e ano de cada edição). Sendo assim, os dados disponíveis na organização a partir de múltiplas fontes são coletados e armazenados.

Para demonstrar a viabilidade do método proposto, um cenário de estudo foi elaborado. O conjunto de dados foi provido pela Base Colaborativa[®], uma organização sem fins lucrativos que arrecada recursos vendendo determinados produtos, entre eles, cursos e viagens de desenvolvimento pessoal. O conjunto de dados conta com 5.046 *leads* e 3 cursos que, juntos somam 50 edições realizadas entre

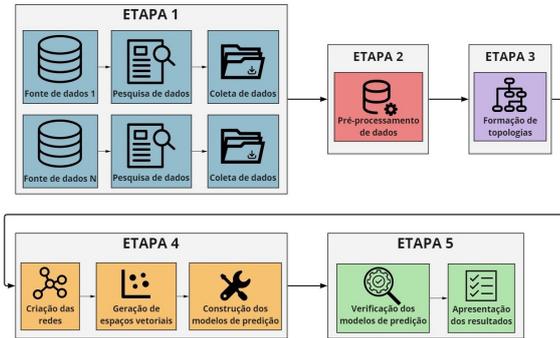


Fig. 1. Visão geral das etapas do método proposto

2016 e 2021. O conjunto de dados inicial foi elaborado a partir de buscas em duas fontes de dados que a organização utiliza: Google Drive[®] e RD Station Marketing[®].

3.2 Etapa 2: Pré-processamento de dados

Nesta segunda etapa, todos os dados coletados são analisados e pré-processados para que apenas as informações úteis à formação de topologias sejam mantidas. Os dados encontravam-se sem uma forma padronizada de estrutura e armazenamento. Ao analisá-los, caso tivessem relação com as edições dos cursos desejados para o estudo, os dados eram inicialmente armazenados em um banco de dados. Após a análise, filtragem e organização, produziu-se uma estrutura composta por diferentes campos, entre eles: *e-mail* do *lead*, edição de interesse do *lead*, ano da edição e curso da edição. Ao todo, foram obtidas 6731 instâncias, onde cada instância representa o registro do interesse de algum *lead* por alguma edição de algum curso.

3.3 Etapa 3: Formação de topologias

Nesta etapa, diversas formações de relações entre os dados pré-processados são elaboradas, resultando em várias possíveis estruturas topológicas. Cada estrutura topológica gera uma rede distinta sobre os mesmos dados. Para a elaboração das topologias não foi utilizado nenhum método específico, mas um processo de abstração e entendimento dos dados coletados e como eles poderiam se relacionar.

Para exemplificar, a Figura 2 ilustra duas topologias diferentes. A Figura 2a relaciona produto com edição do produto e interesse do *lead* com alguma edição. Por outro lado, a topologia da Figura 2b estabelece relações entre as próprias edições de um produto e o interesse de um *lead* com alguma edição. De modo geral, a caracterização de cada vértice depende do estudo que será realizado no contexto da gestão de *leads*.

Depois dos dados coletados e pré-processados, foram criadas 5 topologias diferentes variando as relações estabelecidas entre as categorias dos dados. Para cada topologia foram elaboradas 3 representações vetoriais, duas utilizando o algoritmo *Node2Vec* e uma utilizando o *Metapath2Vec*. Na sequência, foram gerados 2 modelos de predição de *links* para cada representação vetorial utilizando os algoritmos de ML *Random Forest* e *Logistic Regression*, por meio da biblioteca Scikit-learn[®]. No total, 30 modelos de predição de *links* foram construídos para serem analisados.

3.4 Etapa 4: Construção dos modelos de predição

Na sequência, esta quarta etapa prepara uma rede para cada estrutura topológica utilizando-se da biblioteca NetworkX[®], desenvolvida na linguagem de programação Python[®]. Isto ocorre selecionando

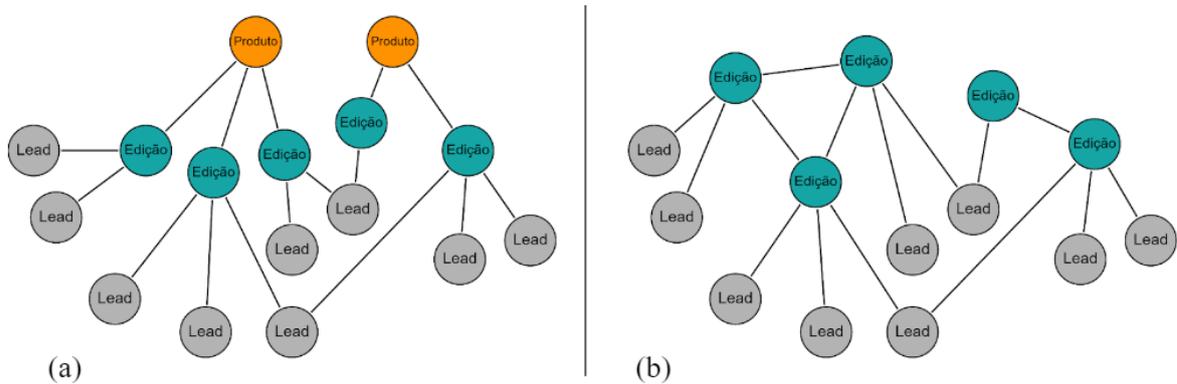


Fig. 2. Ilustração de diferentes topologias

a edição mais recente entre todos os produtos presentes em cada rede e removendo todas as conexões com qualquer *lead*. Os *leads* que foram desvinculados da edição mais recente e que permanecem na rede devido à conexão com alguma outra edição são identificados para serem utilizados na etapa de verificação.

Para realizar o aprendizado de NE, ou seja, as representações vetoriais dos vértices da rede de cada topologia, foram aplicados os algoritmos *Node2Vec* e *Metapath2Vec* com a finalidade de comparação entre os mesmos. Para o *Node2Vec* foi utilizada a biblioteca de mesmo nome do algoritmo, enquanto que para o *Metapath2Vec* foi utilizado a biblioteca StellarGraph[®]. Através do operador binário *Weighted-L1*, foram geradas as representações vetoriais das arestas.

Os modelos de predição de *links* são gerados com algoritmos de ML suportados por vetores de representação das conexões (arestas). Para cada rede, formou-se um conjunto de amostras positivas e negativas. Todas as conexões presentes na rede foram utilizadas no conjunto de amostras positivas. Por outro lado, para o conjunto de amostras negativas, de maneira aleatória, foram selecionadas conexões não existentes na rede. Para não enviesar os modelos, nenhuma amostra negativa continha conexões com a edição mais recente. Com as amostras, os modelos de ML foram treinados para serem utilizados na etapa de verificação.

3.5 Etapa 5: Verificação e apresentação dos resultados

Por fim, nesta última etapa, os vetores que representam conexões entre cada *lead* e a edição mais recente selecionada na Etapa 4 são passados para os modelos de ML treinados, com o intuito de promover um entendimento dessas conexões. Como resultado, obtém-se as probabilidades das conexões entre cada *lead* e a edição selecionada. A partir disso, são verificadas as conexões entre a edição e seus verdadeiros *leads* identificados na Etapa 4 para a obtenção da acurácia de cada modelo de predição.

4. ANÁLISE E DISCUSSÃO DOS RESULTADOS

Através do cenário de estudo, esta seção discute o desempenho dos modelos de predição e a viabilidade de aplicação do método proposto.

A métrica empregada para a avaliação geral do modelo de predição foi a Área Sob a Curva de Característica de Operação do Receptor (do inglês *Area Under the Receiver Operating Characteristic Curve* - ROC-AUC). Esta métrica é amplamente utilizada como indicador de desempenho para problemas de predição de *links* e classificação binária [Lee et al. 2021]. Analisando os valores da ROC-AUC obtido para todos os modelos de predição treinados, percebe-se que estes se comportam de maneira desejada, considerando que o menor valor de ROC-AUC obtido foi de 0,873 e o maior foi de 0,998.

Após a avaliação inicial dos modelos de predição, realizou-se a avaliação do método proposto a partir do cenário de estudo, com o objetivo de determinar seu desempenho real. Analisando os 30 modelos construídos percebe-se que a topologia, a representação vetorial e o algoritmo de ML influenciam significativamente nos resultados dos modelos de predição. Para evidenciar a influência destes componentes, a Tabela 1 apresenta os resultados da representação vetorial 2 para as 5 topologias, uma vez que os resultados obtidos nesta representação estão as maiores acurácias. As últimas 4 colunas das tabelas são explicadas a seguir:

- Predições corretas: considera os 21 *leads* identificados na Etapa 4, como sendo as predições de conexões esperadas para a edição mais recente;
- Tx. de predições corretas: porcentagem de predições corretas referente aos 21 *leads*, $\frac{\text{Predições corretas}}{21}$;
- Todas as predições: apresenta o número total de predições entre todos os *leads* (5018 *leads*) e a edição mais recente;
- Tx. de todas as predições: porcentagem de todas as predições referente aos 5018 *leads*, $\frac{\text{Todas as predições}}{5018}$.

Analisando a Tabela 1 nota-se que as topologias 1 e 5 apresentaram uma baixa acurácia. Na topologia 2 houve uma variação significativa na taxa de predições corretas entre os algoritmos de ML, enquanto que, nas topologias 3 e 4, as taxas de predições corretas ficaram próximas de 100%.

Representação vetorial	Topologia	Algoritmo de ML	Predições corretas	Tx. de predições corretas	Todas as predições	Tx. de todas as predições
Representação vetorial 2	Topologia 1	Random Forest	1	4.76%	16	0.32%
		Logistic Regression	2	9.52%	266	5.30%
	Topologia 2	Random Forest	1	4.76%	489	9.74%
		Logistic Regression	18	85.71%	3998	79.67%
	Topologia 3	Random Forest	21	100.00%	4878	97.21%
		Logistic Regression	20	95.24%	4846	96.57%
	Topologia 4	Random Forest	18	85.71%	4360	86.89%
		Logistic Regression	21	100.00%	4920	98.05%
	Topologia 5	Random Forest	1	4.76%	122	2.43%
		Logistic Regression	2	9.52%	1495	29.79%

Table I. Todos os modelos gerados com a representação vetorial 2

Na Tabela 2, que isola os resultados da topologia 5, fica evidente como a taxa de predições corretas varia de acordo com a representação vetorial e o algoritmo de ML. Porém, a taxa de todas as predições deve também ser considerada para qualificar os modelos. Numa análise ampla, percebe-se que, quanto maior a taxa de predições corretas, maior a taxa de todas as predições, sinalizando um comportamento generalista do modelo. Sendo assim, para qualificar os modelos, utilizou-se a acurácia geral definida na Equação 1, sendo que, quanto mais próximo de 1 for o valor, menos generalista e mais qualificado é o modelo.

$$\text{Acurácia geral} = \frac{\text{Predições corretas}}{\text{Todas as predições}} \quad (1)$$

Desta forma, os 3 melhores resultados considerando a acurácia geral constam na Tabela 3 em ordem decrescente. Nota-se que mesmo os melhores modelos contêm valores muito abaixo de 1, considerado o valor ideal. Assim, levando-se em conta a aplicação do método com os dados disponibilizados no cenário de estudo, entende-se que os modelos gerados tiveram um comportamento muito generalista, com uma elevada taxa de erro nas predições de *links*. Apesar do conjunto de fatores (qualidade dos

Topologia	Representação vetorial	Algoritmo de ML	Predições corretas	Tx. de predições corretas	Todas as predições	Tx. de todas as predições
	Representação vetorial 1	Random Forest	0	0.00%	144	2.87%
		Logistic Regression	11	52.38%	2903	57.85%
Topologia 5	Representação vetorial 2	Random Forest	1	4.76%	122	2.43%
		Logistic Regression	2	9.52%	1495	29.79%
	Representação vetorial 3	Random Forest	6	28.57%	526	10.48%
		Logistic Regression	7	33.33%	383	7.63%

Table II. Todos os modelos gerados com a topologia 5

dados, topologias empregadas e algoritmos de ML utilizados) terem influenciado na acurácia, um fator que se destaca na discussão é a complexidade do aprendizado das representações vetoriais das redes. Quando as redes possuem estruturas heterogêneas e comportamento dinâmico (se modificam com o tempo), existe uma dificuldade maior para aprender representações vetoriais que considerem as informações topológicas e de semântica de forma adequada.

Topologia	Representação vetorial	Algoritmo de ML	Predições corretas	Tx. de predições corretas	Todas as predições	Tx. de todas as predições	Acurácia geral
Topologia 1	Representação vetorial 2	Random Forest	1	4.76%	16	0.32%	0.062
Topologia 5	Representação vetorial 3	Logistic Regression	7	33.33%	383	7.63%	0.018
Topologia 5	Representação vetorial 3	Random Forest	6	28.57%	526	10.48%	0.011

Table III. Melhores modelos de predição

5. CONCLUSÕES E TRABALHOS FUTUROS

Através de várias formações topológicas, este trabalho empregou aprendizado de representação de rede, baseado em RNNs e predições de *links*, para promover suporte à gestão de *leads*. Para tal, efetuou-se uma análise a partir de diversos modelos gerados com diferentes configurações e algoritmos.

Os resultados demonstraram que, tanto a estrutura topológica quanto a representação vetorial e os algoritmos de ML utilizados em cada modelo, influenciam na qualidade dos mesmos. Apesar dos resultados expressivos de ROC-AUC obtidos, grande parte dos modelos se mostraram muito generalistas. Desta forma, utilizou-se a métrica de acurácia geral para auxiliar na qualificação dos modelos, e os três melhores foram discutidos com valores de 0,062, 0,018 e 0,011.

Apesar das contribuições do método proposto, este estudo está sujeito a limitações que requerem trabalhos futuros. Primeiro, com o intuito de lidar com a complexidade de redes heterogêneas e dinâmicas, é imprescindível aprofundar a capacidade do aprendizado de representação vetorial de rede, visando minimizar a perda de informações topológicas e semânticas. Segundo, a falta de padronização e organização das bases de dados do cenário de estudo interferiu diretamente nos resultados do cenário de estudo, limitando a exploração de estruturas topológicas.

Ademais, pesquisas futuras podem explorar outros métodos de representação vetorial com características heterogêneas e dinâmicas das redes. Caberia ainda, investigar o método na prática, ou seja, efetivar uma abordagem de vendas com base nos modelos e, a partir disso, analisar a conversão de *leads* em clientes.

REFERENCES

- AMARA, A., TAIEB, M. A. H., AND AOUICHA, M. B. Network representation learning systematic review: Ancestors and current development state. *Machine Learning with Applications* vol. 6, pp. 100130, 2021.
- BARABASI, A.-L. AND POSFAI, M. Network science, 2016.
- DHARWAL, R. AND KAUR, L. Applications of artificial neural networks: a review. *Indian J. Sci. Technol* 9 (47): 1–8, 2016.
- DONG, Y., CHAWLA, N. V., AND SWAMI, A. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. pp. 135–144, 2017.
- D’HAEN, J., VAN DEN POEL, D., AND THORLEUCHTER, D. Predicting customer profitability during acquisition: Finding the optimal combination of data source and data mining technique. *Expert systems with applications* 40 (6): 2007–2012, 2013.
- D’HAEN, J., VAN DEN POEL, D., THORLEUCHTER, D., AND BENOIT, D. F. Integrating expert knowledge and multilingual web crawling data in a lead qualification system. *Decision Support Systems* vol. 82, pp. 69–78, 2016.
- GEBERT, H., GEIB, M., KOLBE, L., AND RIEMPP, G. Towards customer knowledge management: Integrating customer relationship management and knowledge management concepts, 2002.
- GIACOSA, E., CULASSO, F., AND CROCCO, E. Customer agility in the modern automotive sector: how lead management shapes agile digital companies. *Technological Forecasting and Social Change* vol. 175, pp. 121362, 2022.
- GROVER, A. AND LESKOVEC, J. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 855–864, 2016.
- JÄRVINEN, J. AND TAIMINEN, H. Harnessing marketing automation for b2b content marketing. *Industrial marketing management* vol. 54, pp. 164–175, 2016.
- LEE, J., KO, N., YOON, J., AND SON, C. An approach for discovering firm-specific technology opportunities: Application of link prediction to f-term networks. *Technological Forecasting and Social Change* vol. 168, pp. 120746, 2021.
- MALEK, M., CHEHREGHANI, M. H., NAZERFARD, E., AND CHEHREGHANI, M. H. Shallow node representation learning using centrality indices. In *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 5209–5214, 2021.
- MARTÍNEZ, V., BERZAL, F., AND CUBERO, J.-C. A survey of link prediction in complex networks. *ACM computing surveys (CSUR)* 49 (4): 1–33, 2016.
- MICHELIS, I. Lead lifecycle management: Building a pipeline that never leaks. Tech. rep., Research Report), Aberdeen Group. Retrieved from <https://www.ontargetpartners.com/>, 2009.
- MOLONTAY, R. AND NAGY, M. Two decades of network science: as seen through the co-authorship network of network scientists. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*. pp. 578–583, 2019.
- MONAT, J. P. Industrial sales lead conversion modeling. *Marketing Intelligence & Planning*, 2011.
- OHIOH, A. A., BENYOUCHEF, M., AND ANDREEV, P. Driving inside sales performance with lead management systems: A conceptual model. *Journal of Information Systems Applied Research* 9 (1): 4, 2016.
- PENG, H., LI, J., YAN, H., GONG, Q., WANG, S., LIU, L., WANG, L., AND REN, X. Dynamic network embedding via incremental skip-gram with negative sampling. *Science China Information Sciences* 63 (10): 1–19, 2020.
- SABNIS, G., CHATTERJEE, S. C., GREWAL, R., AND LILIEN, G. L. The sales lead black hole: On sales reps’ follow-up of marketing leads. *Journal of marketing* 77 (1): 52–67, 2013.
- VESPIGNANI, A. AND CALDARELLI, G. *Large scale structure and dynamics of complex networks: from information technology to finance and natural science*. Vol. 2. World scientific, 2007.
- WU, Y.-C. AND FENG, J.-W. Development and application of artificial neural network. *Wireless Personal Communications* 102 (2): 1645–1656, 2018.
- YU, Y.-P. AND CAI, S.-Q. A new approach to customer targeting under conditions of information shortage. *Marketing intelligence & planning*, 2007.