

Relação IA-sociedade representada na mídia: Construção e análise multidisciplinar de redes de entidades nomeadas a partir de textos de notícias

Pedro Henrique de Carvalho Godoy¹, Ruben Interian¹

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
CEP 13.083-970 – Campinas – SP – Brazil

p291115@dac.unicamp.br, ruben@ic.unicamp.br

Abstract. *Artificial Intelligence (AI) is impacting society as a whole. This work creates and analyzes the network of connections between entities present in Brazilian news about AI, examining the most relevant entities to understand how this technology is publicly perceived. Named entity recognition and graph theory were applied to model the network and analyze it from a multidisciplinary perspective. The degree distribution of the vertices suggests a scale-free network structure, indicating the existence of a cumulative advantage effect in the network of entities. The findings demonstrate the effectiveness of interdisciplinary methods in interpreting the structure of the AI debate, highlighting how AI is perceived and discussed publicly.*

Resumo. *A Inteligência Artificial (IA) está impactando toda a sociedade. Este trabalho constrói e analisa a rede de vínculos entre entidades presentes em notícias brasileiras sobre IA, examinando as entidades mais relevantes, para compreender como essa tecnologia é percebida publicamente. Foram aplicadas técnicas de reconhecimento de entidades nomeadas e teoria de grafos para modelar a rede e analisá-la de forma multidisciplinar. A distribuição dos graus dos vértices sugere que a rede é livre de escala, indicando a existência de um efeito de vantagem acumulada na rede de entidades. Os achados demonstram a eficácia de métodos multidisciplinares para interpretar a estrutura do debate sobre a IA, evidenciando como ela é percebida e discutida publicamente.*

1. Introdução

A Inteligência Artificial (IA) é reconhecida como uma Tecnologia de Propósito Geral (TPG) [Kaufman 2022, Brynjolfsson and McAfee 2014], capaz de transformar estruturas sociais e econômicas. Sua aplicação abrange desde análise preditiva criminal [Medium 2023] até assistentes virtuais e terapeutas digitais [Luka 2024, Wisa 2024]. Estudos recentes indicam que a IA pode desempenhar um papel central na mediação de deliberações democráticas [Interian and Rodrigues 2023, Tessler 2024] e na formação de normas sociais [Ikegami 2024]. Diante disso, reconhecer os riscos e benefícios da IA é essencial para garantir um desenvolvimento social equilibrado. [Filippucci et al. 2024].

Identificar quem ou o que está no centro do debate sobre IA é um ponto de partida relevante para compreender os interesses dominantes, as tendências e os possíveis riscos na forma como essa tecnologia é percebida e discutida publicamente. De acordo com Peter Berger e Thomas Luckmann [Berger 1987], a maneira como entendemos o mundo

não é apenas individual, mas moldada coletivamente por meio de interações sociais e discursos legitimados. Isso significa que as narrativas construídas por empresas, governos e mídia sobre a IA, influenciam como a sociedade enxerga e responde a essa tecnologia, determinando desde sua aceitação até suas regulamentações. Da mesma forma, o sociólogo Herbert José de Souza [Souza 1984] reforça que a identificação dos agentes centrais em um debate é essencial para revelar àqueles que orientam a narrativa e influenciam a percepção pública.

Nesta pesquisa, são utilizadas as notícias brasileiras sobre IA como fonte de dados, identificando nelas as entidades e seus vínculos. As notícias atuam como um vetor crucial de disseminação sobre temas emergentes [Mussumeci and Coelho 2018].

Desta forma, o objetivo geral é criar, analisar e caracterizar a estrutura da rede de vínculos entre entidades presentes em notícias brasileiras sobre IA, identificando e examinando as entidades mais relevantes ou centrais nessa rede. Utilizamos técnicas de *Web scraping* para coletar notícias de grandes portais brasileiros e Reconhecimento de Entidades Nomeadas (NER) para identificar termos relevantes nos títulos e textos das publicações, extraíndo vínculos entre as entidades. A rede de entidades é representada por um grafo ponderado, onde as entidades são os nós e os vínculos são as arestas. Aplicamos técnicas de análise de redes para identificar aquelas entidades que estão no centro do debate público sobre IA. Utilizamos teorias e conceitos sociológicos para investigar as relações de força e os aspectos estruturais e conjunturais das entidades e suas conexões.

Entidades nomeadas são termos relevantes encontrados em textos, tais como nomes de pessoas, organizações, lugares e produtos. A técnica NER é uma das abordagens do Processamento de Linguagem Natural (PLN) que identifica e classifica automaticamente estes termos [Zhong and Chen 2021]. Os vínculos entre entidades representam conexões semânticas entre elementos identificados no texto. No contexto deste trabalho, as entidades podem ser empresas, governos, territórios, pessoas e outras palavras representativas, que refletem o debate midiático sobre a IA.

O avanço acelerado da IA exige estudos mais aprofundados e interdisciplinares sobre seus impactos na sociedade. Este trabalho se destaca por adotar a análise de notícias como meio representativo para investigar a relação entre a IA e a sociedade.

2. Captura de Dados e Construção da Rede

Foram coletadas 10.967 notícias em outubro a novembro de 2024 de três grandes portais nacionais (Folha de São Paulo, CNN Brasil e G1), utilizando técnicas de *Web Scraping*. O Reconhecimento de Entidades Nomeadas foi realizado por meio de um modelo de redes neurais convolucionais da biblioteca SpaCy [SpaCy 2025], aplicando a técnica de *fine-tuning* para aprimorar a identificação de entidades do contexto desta pesquisa. Para o processo de *fine-tuning*, criamos uma base com 3 mil notícias sintéticas, geradas automaticamente por meio de um processo que combina aleatoriamente nomes de empresas (130), pessoas (57), tecnologias e algoritmos relacionados à IA (79). As notícias sintéticas seguiram padrões jornalísticos para contextualizar as entidades de forma realista.

Considera-se que entidades extraídas da mesma notícia possuem um vínculo, e o peso deste vínculo é determinado pelo número de vezes que essa coocorrência se repete ao longo do corpus analisado. A rede construída apresenta 7.924 nós e 79.034 arestas, e foi armazenada usando o banco de dados de grafos Neo4j.

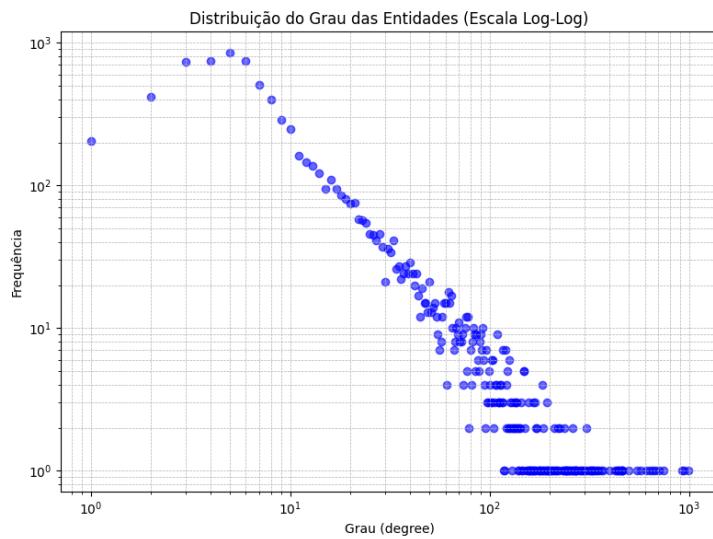


Figura 1. Distribuição dos graus dos vértices na rede de entidades.

3. Análise da Rede e Resultados

Esta seção apresenta os resultados obtidos, destacando padrões de centralidade e interconectividade da rede.

A análise revela uma estrutura altamente assimétrica, com distribuição de graus de cauda longa, sinalizando um típico padrão de rede livre de escala mostrado na Figura 1. Apenas 5% das entidades possuem um grau acima de 81, exercendo influência significativa na propagação de informações na rede. Testes de resiliência confirmam que a rede tolera falhas aleatórias, mas é vulnerável a ataques direcionados, onde a remoção de entidades centrais fragmenta consideravelmente sua estrutura.

Os achados reforçam as propriedades típicas de redes livres de escala, caracterizadas por uma distribuição de lei de potência, onde a maioria dos nós tem poucas ligações, enquanto alguns nós específicos têm um grande número de ligações, conforme a Figura 1.

Foram identificadas 28 comunidades por meio do algoritmo criado por Blondel et al., também conhecido como método de Louvain [Blondel et al. 2008], com modularidade de 0,30, indicando segmentação moderada no debate midiático. A centralidade de intermediação (*betweenness*) revelou que os principais intermediadores incluem veículos de mídia como VEJA e CNN, além de grandes empresas tecnológicas, governos e países como Brasil, EUA e China, conforme a Figura 2.

A distribuição dos pesos das arestas sugere que a cobertura midiática enfatiza questões tecnológicas, econômicas e geopolíticas, refletindo a influência de poucos atores na construção das narrativas sobre IA. Essas evidências corroboram o fenômeno de vantagem acumulada, no qual certas entidades ampliam sua centralidade ao longo do tempo, consolidando seu papel no debate sobre a tecnologia.

4. Discussão e Conclusões

O fenômeno da vantagem acumulada identificado, uma das principais contribuições deste trabalho, revela que certas entidades não apenas influenciam a narrativa, mas ampliam

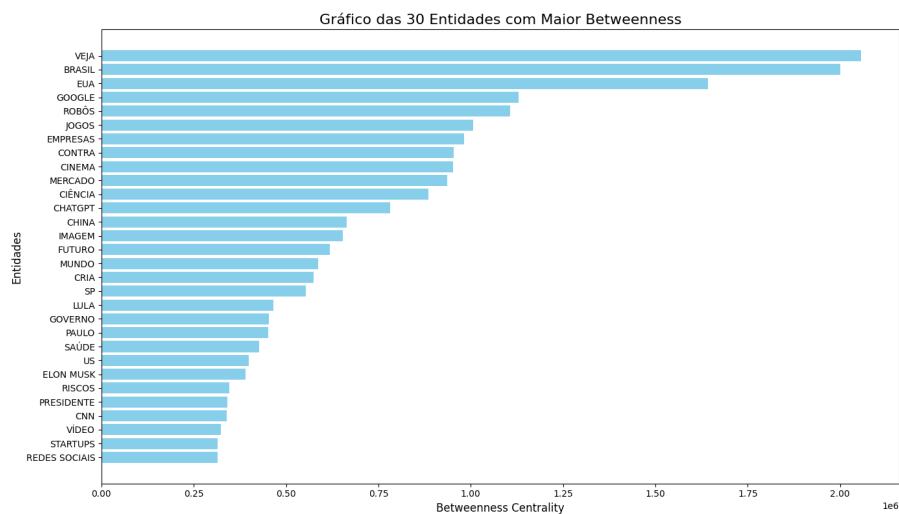


Figura 2. Gráfico das 30 entidades com os maiores valores de Betweenness

progressivamente sua centralidade, consolidando posições dominantes dentro do ecossistema informacional, e indicando que a IA não é apenas um tema emergente, mas um campo discursivo estruturado por relações de poder.

Os resultados evidenciam que o debate midiático sobre a IA é estruturado por um conjunto pequeno de entidades que exercem grande influência sobre a narrativa pública. A nossa análise revelou uma estrutura típica de redes livres de escala, onde poucos nós altamente conectados possuem uma grande influência na rede. O fluxo informacional sobre IA não é homogêneo, sendo direcionado por agentes centrais que moldam a percepção social e a agenda pública sobre a tecnologia. A análise de centralidade identificou que veículos de mídia, grandes empresas e governos desempenham papéis estratégicos na intermediação da informação, configurando um ecossistema no qual determinados atores concentram a construção das narrativas sobre IA.

Entre as limitações do trabalho, destaca-se a seleção de poucas fontes de notícias (três) e o curto tempo de coleta de dados (dois meses). O estudo fornece subsídios para futuras pesquisas sobre a construção discursiva da IA, propondo um pipeline replicável de análise de notícias. Os achados podem orientar a formulação de políticas públicas que promovam um alinhamento entre a evolução tecnológica e interesses sociais, maior pluralidade e transparência no debate sobre IA.

Em trabalhos futuros, pretendemos avaliar métricas adicionais que possam aprofundar a compreensão da rede construída, como a distância média entre as entidades e o diâmetro do grafo, que indica a maior distância entre duas entidades na rede.

Agradecimentos O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil. Processo nº 2024/12936-5. Ruben Interian recebeu apoio do programa PIND da Universidade Estadual de Campinas. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

- [Berger 1987] Berger, T. L. P. L. (1987). *A Construção Social da Realidade - 7^a edição. Tradução de Floriano de Souza Fernandes*. Editora Vozes, Petrópolis-RJ.
- [Blondel et al. 2008] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.
- [Brynjolfsson and McAfee 2014] Brynjolfsson, E. and McAfee, A. (2014). *The Second Machine Age*. W. W. Norton & Company.
- [Filippucci et al. 2024] Filippucci, F., Gal, P., Jona-Lasinio, C., Leandro, A., and Nicoletti, G. (2024). The impact of artificial intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges. *OECD publishing*, 15:1–64.
- [Ikegami 2024] Ikegami, I. H. T. Y. T. (2024). Evolution of social norms in LLM agents using natural language. *arXiv:2409.00993*. <https://arxiv.org/pdf/2409.00993v1.pdf>.
- [Interian and Rodrigues 2023] Interian, R. and Rodrigues, F. A. (2023). Group polarization, influence, and domination in online interaction networks: a case study of the 2022 brazilian elections. *Journal of Physics: Complexity*, 4:035008.
- [Kaufman 2022] Kaufman, D. (2022). *Desmistificando a inteligência artificial*. Autêntica Editora.
- [Luka 2024] Luka (2024). Replika. <https://replika.com>. Acesso em: 01 out. 2024.
- [Medium 2023] Medium, P. (2023). Compas — an AI tool sending or keeping people in jail. <https://fatiLexica.medium.com/compas-an-ai-tool-sending-or-keeping-people-in-jail-d9228df3a2c6>. Acesso em: 01 out. 2024.
- [Mussumeci and Coelho 2018] Mussumeci, E. and Coelho, F. C. (2018). Reconstructing news spread networks and studying its dynamics. *Social Network Analysis and Mining*.
- [Souza 1984] Souza, H. J. (1984). *Como se faz Análise de Conjuntura*. Editora Vozes, Petrópolis-RJ, 16^º edition.
- [SpaCy 2025] SpaCy (2025). Available trained pipelines for portuguese. Disponível em: <https://spacy.io/models/pt>. Acesso em: 9 abr. 2025.
- [Tessler 2024] Tessler, M. H. (2024). AI can help humans find common ground in democratic deliberation. *Science*, 386:6719.
- [Wisa 2024] Wisa (2024). Wisa. Disponível em: <https://www.wysa.com>. Acesso em: 01 out. 2024.
- [Zhong and Chen 2021] Zhong, Z. and Chen, D. (2021). A frustratingly easy approach for entity and relation extraction. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.