



**UFOP**

Universidade Federal  
de Ouro Preto

**Universidade Federal de Ouro Preto  
Instituto de Ciências Exatas e Aplicadas  
Departamento de Computação e Sistemas**

**Análise de Redes Complexas para  
Modelagem de Relações Competitivas  
no Esporte Olímpico**

**Caio Damasceno Alves**

**TRABALHO DE  
CONCLUSÃO DE CURSO**

**ORIENTAÇÃO:**

Helen de Cassia Sousa da Costa Lima

**Março, 2025**

**João Monlevade–MG**

**Caio Damasceno Alves**

# **Análise de Redes Complexas para Modelagem de Relações Competitivas no Esporte Olímpico**

Orientador: Helen de Cassia Sousa da Costa Lima

Monografia apresentada ao curso de Sistemas de Informação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

**Universidade Federal de Ouro Preto**

**João Monlevade**

**Março de 2025**

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

A474a Alves, Caio Damasceno.  
Análise de redes complexas para modelagem de relações  
competitivas no esporte olímpico. [manuscrito] / Caio Damasceno Alves. -  
2026.

102 f.: il.: color., gráf., tab..

Orientadora: Profa. Dra. Helen Lima.  
Monografia (Bacharelado). Universidade Federal de Ouro Preto.  
Instituto de Ciências Exatas e Aplicadas. Graduação em Sistemas de  
Informação .

1. Grafos. 2. Esportes - Desempenho. 3. Modelos matemáticos. 4.  
Olimpíadas - Esportes. I. Lima, Helen. II. Universidade Federal de Ouro  
Preto. III. Título.

CDU 519.17:796.032.2

Bibliotecário(a) Responsável: Flavia Reis - CRB6/2431



## FOLHA DE APROVAÇÃO

**Caio Damasceno Alves**

### **Análise de Redes Complexas para Modelagem de Relações Competitivas no Esporte Olímpico**

Monografia apresentada ao Curso de Sistemas de Informação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação

Aprovada em 04 de Março de 2026

#### Membros da banca

Dra. Helen de Cássia Sousa da Costa Lima - Orientadora (Universidade Federal de Ouro Preto)  
Dra. Gilda Aparecida de Assis (Universidade Federal de Ouro Preto)  
Dra. Janniele Aparecida Soares Araújo (Universidade Federal de Ouro Preto)

Helen de Cássia Sousa da Costa Lima, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 20/03/2026



Documento assinado eletronicamente por **Helen de Cassia Sousa da Costa Lima, PROFESSOR DE MAGISTERIO SUPERIOR**, em 20/03/2026, às 16:29, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1079203** e o código CRC **E69BF6A8**.

*À minha família, pelo apoio incondicional em cada etapa desta jornada.*

# Agradecimentos

Agradeço primeiramente à minha família, pelo amor, paciência e suporte ao longo de toda a minha formação acadêmica.

À minha orientadora, Prof.<sup>a</sup> Helen de Cassia Sousa da Costa Lima, pela orientação dedicada, pelos ensinamentos e pela confiança depositada neste trabalho.

Aos professores do Departamento de Computação e Sistemas da Universidade Federal de Ouro Preto, que contribuíram para a minha formação e ampliaram minha visão sobre a ciência e a tecnologia.

Aos meus amigos e colegas de curso, pelas trocas de experiências, pelo companheirismo e por tornarem essa trajetória mais leve e significativa.

E a todos que, de alguma forma, contribuíram para a realização deste trabalho.

*“Science is more than a body of knowledge; it is a way of thinking.”*

— Carl Sagan (1934 – 1996),  
*in: The Demon-Haunted World: Science as a Candle in the Dark.*

# Resumo

Esta monografia apresenta a aplicação de análise de redes complexas a dados históricos olímpicos (1896–2016), modelando competições esportivas através de grafos direcionados e ponderados que revelam estruturas competitivas latentes. O trabalho analisa 9.510 atletas medalhistas em seis modalidades (Swimming, Basketball, Football, Athletics, Judo e Boxing) ao longo de 120 anos de história olímpica. As redes construídas representam atletas como nós e relações competitivas baseadas em confrontos diretos no pódio como arestas ponderadas por hierarquia de medalhas. Através de PageRank adaptado ao contexto esportivo, métricas de centralidade e detecção de comunidades via algoritmo de Louvain, o estudo identifica hierarquias de performance, agrupamentos estruturais e padrões temporais não evidentes em análises tradicionais baseadas em contagem de medalhas. Os resultados revelam diferenças sistemáticas entre modalidades: esportes individuais produzem redes altamente modulares com comunidades segregadas, enquanto esportes coletivos geram estruturas densamente interconectadas com baixa modularidade. A caracterização multidimensional de comunidades através de entropia temporal, concentração geográfica e índices de dominância permite compreensão profunda das dinâmicas competitivas olímpicas. Um dashboard interativo desenvolvido em Streamlit facilita a exploração dos resultados e a validação dos achados.

**Palavras-chaves:** redes complexas, esporte olímpico, PageRank, detecção de comunidades, análise de desempenho esportivo.

# Abstract

This monograph presents the application of complex network analysis to historical Olympic data (1896–2016), modeling sports competitions through directed and weighted graphs that reveal latent competitive structures. The work analyzes 9,510 medalist athletes across six sports (Swimming, Basketball, Football, Athletics, Judo, and Boxing) spanning 120 years of Olympic history. Networks represent athletes as nodes and competitive relationships based on shared podium confrontations as edges weighted by medal hierarchy. Through PageRank adapted to sports contexts, centrality metrics, and community detection via Louvain algorithm, the study identifies performance hierarchies, structural clusters, and temporal patterns not evident in traditional medal-count analyses. Results reveal systematic differences between modalities: individual sports produce highly modular networks with segregated communities, while team sports generate densely interconnected structures with low modularity. Multidimensional community characterization through temporal entropy, geographic concentration, and dominance indices enables deep understanding of Olympic competitive dynamics. An interactive Streamlit dashboard facilitates results exploration and findings validation.

**Keywords:** complex networks, Olympic sport, PageRank, community detection, sports performance analysis.

# Lista de ilustrações

- Figura 1 – Fluxo metodológico completo: da aquisição dos dados históricos olímpicos à geração de resultados e ao *dashboard* interativo. . . . . 38
- Figura 2 – Comparação de densidade entre eventos masculinos e femininos. Cinco dos seis esportes apresentam densidade feminina superior: Boxe exibe razão F/M de 25.65×, Futebol 5.52×, Basquete 1.89×, Judô 2.66× e Atletismo 100m 1.30×. Apenas Natação 100m Nado Livre apresenta densidade masculina ligeiramente superior (razão 0.90×). Maior densidade feminina sugere campos competitivos menores com recorrência mais intensa entre mesmas atletas. . . . . 62
- Figura 3 – Top 3 atletas por *PageRank* em cada esporte, segregados por gênero (M: masculino, F: feminino). Rankings são independentes: valores não são comparáveis entre esportes devido a diferenças estruturais (coletivo vs individual, tamanho de rede, topologia competitiva). Atletas destacados incluem: Usain Bolt e Elaine Thompson (Atletismo), Carmelo Anthony e Teresa Edwards (Basquete), Dawn Fraser e Johnny Weissmuller (Natação), entre outros. . . . . 67
- Figura 4 – Índice de segregação estrutural por evento. Esportes individuais apresentam segregação completa (índice 52-100): Atletismo Feminino 100m, Judô Masculino -90kg e Atletismo Masculino 100m exibem comunidades totalmente delimitadas. Esportes coletivos mostram baixa segregação (1.5-15.8), indicando fronteiras difusas entre agrupamentos temporais. . . . . 68
- Figura 5 – Evolução do percentual de medalhas dos Top 3 países por era olímpica. Era Pioneira (1896-1936) apresenta concentração de 74.0%, seguida de democratização na Era Moderna (59.6% em 1984-2000). Era Contemporânea (2004-2016) reverte o padrão, elevando concentração para 64.4%, evidenciando re-centralização competitiva apesar de globalização crescente. . . . . 69
- Figura 6 – Impacto de remoção de 10% dos nós: comparação entre remoção aleatória e remoção direcionada por *PageRank*. Remoção aleatória causa fragmentação mínima (+0.2 componentes em média), enquanto remoção direcionada fragmenta severamente (+7.9 componentes), evidenciando vulnerabilidade assimétrica típica de redes *scale-free* com *hubs* críticos. 70

Figura 7 – Evolução temporal da densidade para as 12 redes olímpicas. Redes masculinas (azul) apresentam trajetórias monotonicamente decrescentes ao longo de períodos históricos extensos (1896-2016). Redes femininas (vermelho) iniciam com densidades elevadas (>40%) e decaem rapidamente nas primeiras décadas, convergindo assintoticamente para níveis similares aos masculinos após 3-4 décadas. . . . .	72
Figura 8 – Evolução temporal da modularidade (Louvain) para as 12 redes. Apesar de declínio sistemático de densidade, modularidade permanece elevada ( $Q > 0.60$ em 95% dos snapshots temporais), evidenciando que estrutura comunitária persiste mesmo em redes crescentemente esparsas. Este padrão indica que competição olímpica organiza-se em clusters temporais/geográficos robustos. . . . .	73
Figura 9 – Comparação estrutural entre redes de Futebol Olímpico masculino e feminino. Nós dimensionados por PageRank, cores representam comunidades detectadas por algoritmo Louvain. Diferença histórica (76 anos) resulta em estruturas radicalmente distintas: masculina fragmentada temporalmente, feminina coesa e densa. . . . .	75
Figura 10 – Redes competitivas de Basquete Olímpico. Cor por comunidade ( <i>Louvain</i> ), tamanho por <i>PageRank</i> , layout <i>ForceAtlas2</i> . . . . .	76
Figura 11 – Comparação estrutural entre redes de Atletismo 100m masculino e feminino. Nós dimensionados por PageRank, cores representam comunidades detectadas. Ambas exibem segregação extrema característica de eventos de velocidade, onde carreiras atléticas são curtas e gerações competitivas raramente se sobrepõem. . . . .	77
Figura 12 – Redes competitivas de Natação 100m Livre Olímpico. Cor por comunidade ( <i>Louvain</i> ), tamanho por <i>PageRank</i> , layout <i>ForceAtlas2</i> . . . . .	78
Figura 13 – Redes competitivas de Judô Olímpico. Cor por comunidade ( <i>Louvain</i> ), tamanho por <i>PageRank</i> , layout <i>ForceAtlas2</i> . . . . .	78
Figura 14 – Redes competitivas de Boxe Olímpico. Cor por comunidade ( <i>Louvain</i> ), tamanho por <i>PageRank</i> , layout <i>ForceAtlas2</i> . . . . .	79
Figura 15 – Diversidade geográfica (entropia) versus concentração de medalhas nos top 3 países. Redes masculinas (azul) e femininas (vermelho) segregam espacialmente: redes femininas concentram-se no quadrante inferior esquerdo (baixa diversidade, alta concentração), enquanto redes masculinas dispersam-se amplamente. Exceção: Futebol Feminino apresenta entropia intermediária apesar de dominância elevada de USA. . . . .	81

Figura 16 – Top 5 países por número de medalhas em cada uma das 12 redes individuais. USA domina 7 de 12 modalidades (Basketball M/F, Athletics M/F, Swimming M/F, Boxing M). Brasil lidera Futebol Masculino, Japão lidera Judô Masculino, França lidera Judô Feminino, e China lidera Boxing Feminino. Redes masculinas (azul) apresentam distribuição mais balanceada entre top 5 países comparado a redes femininas (vermelho).	82
Figura 17 – Coeficiente de Gini por rede individual, ordenado por magnitude. Athletics 100m Masculino apresenta maior desigualdade (Gini = 0.626), onde USA concentra quase metade de todas as medalhas históricas. Judô Feminino -70kg apresenta menor desigualdade (Gini = 0.150), com distribuição mais equitativa entre 12 países participantes. Linha tracejada indica <i>threshold</i> de desigualdade moderada (Gini = 0.50).	83
Figura 18 – Distribuições de métricas estruturais com identificação de outliers via IQR. Redes masculinas (azul) e femininas (vermelho) exibem dispersão assimétrica: densidade feminina concentra-se em valores mais altos, enquanto número de nós apresenta distribuição log-normal com Futebol Masculino como outlier extremo superior.	85
Figura 19 – Densidade versus modularidade com rótulos individuais das 12 redes. Boxe Médio F (densidade 0.417) posiciona-se como outlier extremo à direita, mas mantém relação estrutural contínua com outras redes de alta densidade (Judô, Basquete F). Futebol M posiciona-se no extremo inferior esquerdo (densidade mínima), mas sem descontinuidade topológica.	86
Figura 20 – Percentual de especialização em medalhas de ouro por nível hierárquico. Comunidades periféricas apresentam 35.42% de especialização, superior ao núcleo (31.37%). Padrão contra-intuitivo sugere que dominância absoluta isola competidores da estrutura central da rede.	87

# Lista de tabelas

Tabela 1 – Estrutura de metadados do dataset olímpico histórico (1896-2016) . . . .	40
Tabela 2 – Hierarquia e dimensões do conjunto de dados analisado . . . . .	42
Tabela 3 – Validação empírica Louvain vs Infomap nas 12 redes principais do estudo	57
Tabela 4 – Características estruturais dos 12 casos de estudo selecionados via <i>Iconic Score</i> . Redes variam substancialmente em tamanho (7-1.275 atletas), densidade (1.33%-23.81%) e segregação estrutural (1.50-100), refletindo diferenças em tipologia esportiva, gênero e longevidade histórica. . . . .	59
Tabela 5 – Análise de sensibilidade de sistemas de ponderação de medalhas . . . . .	61
Tabela 6 – Validação estatística: densidade feminina vs masculina . . . . .	63
Tabela 7 – Métricas agregadas com intervalos de confiança (IC95%) via bootstrap	64
Tabela 8 – <i>Dataset</i> completo das 12 redes para análise de regressão . . . . .	65
Tabela 9 – Análise de regressão: efeito do gênero na densidade da rede controlando para confounders . . . . .	65
Tabela 10 – Pontos de mudança estrutural identificados via análise temporal dinâmica	71
Tabela 11 – Dominância geográfica e diversidade nas 12 redes individuais . . . . .	80
Tabela 12 – Outliers identificados via método IQR (Interquartile Range) . . . . .	85

# Lista de símbolos

$G = (V, E)$	Grafo composto por conjunto de vértices $V$ e arestas $E$
$w(e)$	Peso da aresta $e$ em um grafo ponderado
$PR(v)$	PageRank do vértice $v$
$d$	Fator de amortecimento do PageRank (tipicamente 0.85)
$k$	Grau de um nó (número de conexões)
$k_{in}$	Grau de entrada (número de arestas entrantes)
$k_{out}$	Grau de saída (número de arestas saíntes)
$Q$	Modularidade da rede
$C_i$	Comunidade $i$ em uma partição da rede
$H$	Entropia de Shannon
$\alpha$	Nível de significância estatística
$\rho$	Coefficiente de correlação

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
1.1	Questão de Pesquisa	18
1.2	Objetivos	18
1.3	Contribuições	19
1.4	Visão Geral da Metodologia	19
1.5	Organização do Trabalho	20
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>21</b>
2.1	Fundamentação Teórica em Redes Complexas	21
2.2	Modelagem de Redes Competitivas Esportivas	22
2.3	Métricas de Centralidade e Ranking em Redes	24
2.4	Detecção e Caracterização de Comunidades	26
2.4.1	Métricas de Desigualdade e Normalização Cross-Network	29
2.5	Validação Estatística em Redes Complexas	30
2.5.1	Controle de Variáveis de Confusão em Comparações Estruturais	31
2.6	Trabalhos Relacionados	31
2.6.1	Síntese Comparativa	33
2.6.2	Lacunas na Literatura	35
2.7	Ferramentas e Tecnologias Utilizadas	36
<b>3</b>	<b>DESENVOLVIMENTO</b>	<b>37</b>
3.1	Seleção de Modalidades Esportivas	37
3.2	Preparação e Curadoria dos Dados	39
3.3	Modelagem de Redes Competitivas	41
3.3.1	Formalização do Iconic Score	43
3.4	Implementação Técnica e Pipeline de Análise	45
3.5	Validação da Escolha de Algoritmo de Detecção de Comunidades	46
3.5.1	Metodologia de Validação	46
3.5.2	Critérios de Decisão	46
3.5.3	Decisão e Justificativa	47
3.6	Métricas de Análise de Redes	47
3.7	Análise de Confounders em Comparações Cross-Gender	49
3.7.1	Identificação de Confounders	49
3.7.2	Especificação do Modelo de Regressão	49
3.7.3	Procedimento de Inferência Estatística	50
3.7.4	Construção do <i>Dataset</i> de Análise	50

<b>3.8</b>	<b>Dashboard Interativo de Visualização</b>	<b>51</b>
3.8.1	Motivação e Objetivos	51
3.8.2	Visualização Interativa de Redes com vis-network	52
3.8.3	Arquitetura de Publicação Web	53
3.8.4	Limitações e Trade-offs	54
<b>4</b>	<b>RESULTADOS</b>	<b>56</b>
<b>4.1</b>	<b>Validação Empírica de Algoritmos: Louvain vs Infomap</b>	<b>56</b>
4.1.1	Concordância Estrutural: Evidência Quantitativa	56
4.1.2	Qualidade das Partições: Modularidade Comparável	57
4.1.3	Padrões Cross-Esporte	57
4.1.4	Síntese da Validação Empírica	58
<b>4.2</b>	<b>Visão Geral dos Casos de Estudo</b>	<b>58</b>
<b>4.3</b>	<b>Análise de Sensibilidade do Sistema de Ponderação</b>	<b>60</b>
4.3.1	Metodologia de Validação	60
4.3.2	Resultados da Validação	60
4.3.3	Implicações Metodológicas	61
<b>4.4</b>	<b>Análise Comparativa Cross-Esporte</b>	<b>62</b>
4.4.1	Diferenças de Densidade por Gênero	62
4.4.2	Análise de Confounders: Controlando Variáveis Estruturais	64
4.4.3	Atletas com Maior Centralidade Estrutural por Esporte	67
4.4.4	Segregação Estrutural por Evento	68
4.4.5	Evolução Temporal da Concentração Geográfica	69
4.4.6	Resiliência e Robustez Estrutural	70
4.4.7	Dinâmica Temporal: Evolução Estrutural por Década	71
<b>4.5</b>	<b>Análise Intra-Esporte: Comparações por Evento Específico</b>	<b>74</b>
4.5.1	Futebol (Futebol)	74
4.5.2	Basquete (Basquete)	75
4.5.3	Atletismo 100m (Atletismo 100 Metros Rasos)	76
4.5.4	Natação 100m Nado Livre (Natação 100 Metros Livre)	76
4.5.5	Judô (Judô)	77
4.5.6	Boxe (Boxe)	79
<b>4.6</b>	<b>Distribuição Geográfica e Padrões de Dominância</b>	<b>79</b>
4.6.1	Diversidade Geográfica: Entropia e Número de Países	80
4.6.2	Concentração de Medalhas: Top 3 Dominância	80
4.6.3	Hegemonia Regional Específica por Esporte	82
4.6.4	Desigualdade Estrutural: Coeficiente de Gini	83
4.6.5	Síntese da Análise Geográfica	84
4.6.6	Validação Estatística: Análise Formal de Outliers	84
<b>4.7</b>	<b>Descobertas Contra-Intuitivas</b>	<b>87</b>

4.7.1	Especialização em Ouro na Periferia . . . . .	87
4.7.2	Redes Femininas Sistemáticamente Mais Densas . . . . .	88
4.7.3	Re-Concentração Geográfica Contemporânea . . . . .	88
<b>4.8</b>	<b>Validação com Estatísticas Históricas Conhecidas . . . . .</b>	<b>88</b>
4.8.1	Futebol: Brasil e a Era Dourada Húngara . . . . .	89
4.8.2	Basquete: Hegemonia Estadounidense e Rivalidade da Guerra Fria . . . . .	89
4.8.3	Atletismo 100m: Confirmação de Hegemonia Estadounidense . . . . .	90
4.8.4	Natação 100m Livre: Atletas Históricos Confirmados . . . . .	90
4.8.5	Judô: Domínio Japonês e Emergência Francesa . . . . .	90
4.8.6	Boxe: Rivalidade USA-Cuba e Introdução Feminina Recente . . . . .	91
4.8.7	Síntese da Validação . . . . .	91
<b>4.9</b>	<b>Síntese dos Resultados . . . . .</b>	<b>92</b>
<b>5</b>	<b>CONCLUSÃO . . . . .</b>	<b>93</b>
<b>5.1</b>	<b>Síntese dos Achados Principais . . . . .</b>	<b>93</b>
<b>5.2</b>	<b>Contribuições Metodológicas . . . . .</b>	<b>94</b>
<b>5.3</b>	<b>Limitações do Estudo . . . . .</b>	<b>95</b>
<b>5.4</b>	<b>Trabalhos Futuros . . . . .</b>	<b>96</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>98</b>

# 1 Introdução

Desde a primeira olimpíada moderna em Atenas 1896, mais de 135 mil atletas participaram dos Jogos Olímpicos ao longo de 120 anos de história (até Rio 2016). Por trás desses resultados individuais, existe uma rede complexa de relações competitivas que revela padrões latentes de dominância, rivalidades e agrupamentos estruturais não capturados por métricas tradicionais. A modelagem e a análise de redes complexas têm sido amplamente aplicadas para compreender sistemas complexos em diversas áreas do conhecimento, revelando padrões estruturais e dinâmicas não capturados por análises tradicionais baseadas em métricas isoladas (AHMED et al., 2023). No contexto esportivo, esta abordagem oferece perspectivas complementares ao permitir a modelagem de competições como sistemas relacionais, onde atletas e suas interações competitivas podem ser representados como grafos direcionados e ponderados (RADICCHI, 2016).

Especificamente, a análise de redes complexas revela propriedades estruturais complementares não capturadas por métricas tradicionais. Enquanto estatísticas convencionais avaliam atletas isoladamente através de tempo, pontos ou recordes, a análise de redes captura hierarquias de influência, padrões de dominância competitiva e agrupamentos estruturais (WÄSCHE et al., 2017; LÓPEZ-FELIP et al., 2018).

Consequentemente, esta metodologia permite compreender o ecossistema competitivo através de interações entre atletas de diferentes modalidades e países ao longo do tempo, considerando competições esportivas como sistemas dinâmicos complexos (DAVIDS et al., 2013).

Embora trabalhos recentes tenham aplicado redes complexas ao contexto esportivo (MOTEGI; MASUDA, 2012; HSU; ZHANG, 2025), a maioria concentra-se em uma análise isolada de modalidades específicas, sem explorar diferenças estruturais sistemáticas entre tipos de interdependência competitiva ao longo de períodos históricos extensos.

Portanto, o presente trabalho preenche esta lacuna aplicando ferramentas de ciência de redes a dados olímpicos históricos (1896-2016), desenvolvendo uma metodologia que integra a modelagem de grafos competitivos por evento específico (*per-event*), a análise de centralidade e a detecção de comunidades. Esta abordagem gerou 143 redes competitivas direcionadas e ponderadas construídas pela chave composta (Ano, Evento\_Normalizado), das quais 12 casos de estudo representativos foram selecionados através de amostragem por critério (PALINKAS et al., 2015) com pontuação multi-dimensional denominada *Iconic Score*. A análise abrange seis modalidades olímpicas que representam dois tipos estruturais de interdependência competitiva: esportes coletivos (Futebol e Basquete), esportes individuais de performance (Natação e Atletismo) e esportes individuais de

combate (Judô e Boxe). Os 12 eventos específicos analisados são: Futebol Masculino, Futebol Feminino, Basquete Masculino, Basquete Feminino, Natação 100m Nado Livre Masculino, Natação 100m Nado Livre Feminino, Atletismo 100m Masculino, Atletismo 100m Feminino, Judô Meio-Médio (-73kg) Masculino, Judô Meio-Médio (-70kg) Feminino, Boxe Peso Médio Masculino, e Boxe Peso Médio Feminino.

Estabelecido o contexto e as lacunas da literatura científica existente, este trabalho busca responder à seguinte questão de pesquisa:

## 1.1 Questão de Pesquisa

Este trabalho investiga a seguinte questão:

*Como a estrutura de redes competitivas olímpicas difere entre esportes com diferentes tipos de interdependência competitiva (individual de performance, individual de combate, e coletivo)?*

Especificamente, busca-se identificar que padrões de dominância e agrupamento estrutural emergem ao longo de mais de um século de competições olímpicas.

## 1.2 Objetivos

Este trabalho objetiva aplicar a análise de redes complexas a dados históricos olímpicos (1896-2016) para identificar e quantificar diferenças estruturais entre modalidades com tipos distintos de interdependência competitiva (individual de performance, individual de combate, e coletivo) através de métricas de centralidade, detecção de comunidades e caracterização multidimensional de agrupamentos estruturais.

Os objetivos específicos compreendem:

- Construir 143 redes competitivas direcionadas e ponderadas aplicando modelagem *per-event* através da chave composta (Ano, `Evento_Normalizado`), onde `Evento_Normalizado` é a coluna do *dataset* que identifica a prova esportiva específica dentro de cada modalidade (e.g., *100m Nado Livre* em Natação, *100m Rasos* em Atletismo), gerando redes competitivas independentes para cada prova realizada em cada edição dos jogos. Das 143 redes geradas, selecionar 12 casos de estudo representativos via critério *Iconic Score*, abrangendo atletas medalhistas nas modalidades de Natação, Atletismo, Basquete, Futebol, Judô e Boxe;
- Aplicar métricas de centralidade (*PageRank*, *degree*, *betweenness*, *closeness*) para identificar atletas estruturalmente importantes nas redes competitivas;

- Detectar e caracterizar comunidades nas redes através do algoritmo de *Lowvain*, analisando suas propriedades temporais, geográficas e de performance;
- Implementar *dashboard* interativo como artefato complementar para exploração e visualização dos resultados;
- Analisar comparativamente as propriedades estruturais das redes entre esportes coletivos, individuais de performance e individuais de combate.

### 1.3 Contribuições

Este trabalho apresenta três contribuições principais para a literatura de análise de redes aplicada ao esporte:

**Caracterização Multidimensional de Comunidades:** A análise de comunidades desenvolvida integra métricas estruturais de tamanho e modularidade com a caracterização temporal através de entropia de Shannon e heterogeneidade de performance através do coeficiente de variação de *PageRank*. Esta sistematização permite compreender não apenas a existência de agrupamentos estruturais, mas suas características distintivas no contexto competitivo de cada modalidade.

**Análise Comparativa Multi-Esporte:** A análise estrutural de competições com diferentes tipos de interdependência (individual de performance, individual de combate, e coletivo) ao longo de 120 anos de história olímpica permite a identificação de padrões sistemáticos relacionados ao formato competitivo.

**Dashboard Interativo para Exploração:** A implementação de uma interface web interativa oferece um artefato complementar que demonstra a viabilidade de ferramentas acessíveis para exploração de redes complexas por audiências não-técnicas. O *dashboard* integra a análise de centralidades, a visualização de comunidades, a análise temporal e comparações entre esportes, facilitando a identificação de padrões e a validação de hipóteses sobre estruturas competitivas.

### 1.4 Visão Geral da Metodologia

A metodologia desenvolvida segue um fluxo de cinco etapas principais, detalhadas ao longo do Capítulo 3:

1. **Curadoria dos dados:** filtragem e preparação do *dataset* histórico olímpico (Griffin 2018), selecionando registros de medalhistas e normalizando eventos para garantir comparabilidade entre edições;

2. **Construção das redes:** criação de 143 redes competitivas direcionadas e ponderadas por chave composta (Ano, Evento\_Normalizado), com arestas orientadas do derrotado para o vencedor e pesos proporcionais à diferença de posição no pódio;
3. **Seleção de casos de estudo:** aplicação do *Iconic Score* para selecionar 12 redes representativas de 6 modalidades, equilibrando cobertura temporal, diversidade geográfica, volume de medalhistas e representação de gênero;
4. **Análise de centralidade:** cálculo de *PageRank*, *betweenness*, *degree* e *closeness* para identificar atletas estruturalmente importantes em cada rede;
5. **Detecção e caracterização de comunidades:** aplicação do algoritmo de *Louvain* com caracterização multidimensional (temporal, geográfica e de performance) dos agrupamentos detectados.

## 1.5 Organização do Trabalho

Este trabalho está organizado em cinco capítulos. O Capítulo 2 apresenta a revisão bibliográfica, abordando os fundamentos teóricos de redes complexas, a modelagem de redes competitivas esportivas, as métricas de centralidade e os algoritmos de detecção de comunidades, além de trabalhos relacionados que aplicaram a análise de redes ao contexto esportivo e as ferramentas e tecnologias utilizadas. O Capítulo 3 apresenta o desenvolvimento do trabalho, incorporando a metodologia adotada e sua implementação técnica: detalhando o processo de coleta e preparação dos dados olímpicos, a construção das redes competitivas, os critérios de ponderação de arestas, e os procedimentos de análise através de métricas de centralidade e detecção de comunidades. O Capítulo 4 apresenta os resultados obtidos, incluindo a caracterização estrutural das redes, os rankings de atletas por *PageRank*, a análise de comunidades detectadas e as comparações entre modalidades esportivas. O Capítulo 5 conclui o trabalho, sintetizando as principais contribuições, limitações e direções para pesquisas futuras.

## 2 Revisão bibliográfica

A análise de redes complexas oferece uma perspectiva complementar para a investigação de sistemas competitivos, permitindo modelar relações entre competidores através de grafos direcionados e ponderados. A abordagem fundamenta-se em um arcabouço teórico consolidado que integra a teoria de grafos, as métricas de centralidade e os algoritmos de detecção de comunidades. Este capítulo revisa os conceitos fundamentais e as metodologias que embasam as escolhas técnicas adotadas neste trabalho.

### 2.1 Fundamentação Teórica em Redes Complexas

A teoria de redes complexas oferece uma forma unificada de compreender sistemas intrincados em múltiplas disciplinas. Modelar entidades como nós conectados por relações específicas revela propriedades emergentes que não aparecem em análises tradicionais baseadas em métricas isoladas (AHMED et al., 2023). Esta abordagem encontrou aplicação em sistemas biológicos, sociais, tecnológicos e econômicos.

As redes podem ser representadas matematicamente como grafos  $G = (V, E)$ , onde  $V$  é o conjunto de vértices (nós) e  $E$  é o conjunto de arestas que conectam pares de vértices. Em grafos direcionados, cada aresta possui uma orientação específica, representando relações assimétricas entre entidades; em grafos ponderados, cada aresta possui um valor numérico quantificando a intensidade da conexão.

No contexto esportivo, esta representação captura hierarquias competitivas com precisão: a direção da aresta indica o resultado (perdedor  $\rightarrow$  vencedor), enquanto o peso reflete a magnitude da derrota ou diferença hierárquica. Grafos direcionados e ponderados capturam tanto informação ordinal (quem derrotou quem) quanto cardinal (com que intensidade), permitindo análises mais ricas que abordagens binárias (LEICHT; NEWMAN, 2008).

Além da representação matemática básica, redes complexas exibem propriedades emergentes que surgem da interação entre seus componentes, não estando presentes em elementos individuais isolados. Três propriedades estruturais fundamentais caracterizam estas redes. Primeiro, a distribuição de grau heterogênea, onde alguns nós possuem muitas conexões enquanto a maioria possui poucas, fenômeno observado em diversas redes naturais e artificiais. Segundo, o coeficiente de agrupamento elevado, que reflete a tendência de vizinhos de um nó estarem conectados entre si, formando triângulos e estruturas localmente densas. Terceiro, a existência de comunidades, que se manifesta através de grupos densamente conectados internamente mas esparsamente conectados

entre si (FORTUNATO, 2010).

Dois mecanismos fundamentais moldam a formação e a evolução de redes complexas. O *preferential attachment* (ligação preferencial) descreve o fenômeno onde nós já altamente conectados têm maior probabilidade de receber novas conexões, seguindo o princípio do “rico fica mais rico” (BARABÁSI; ALBERT, 1999). Em redes esportivas, este mecanismo manifesta-se quando atletas de elite enfrentam continuamente outros competidores de alto nível, acumulando confrontos de qualidade que amplificam sua centralidade estrutural. A *homofilia* refere-se à tendência de nós com características similares estabelecerem conexões preferencialmente entre si, capturada pelo princípio sociológico de que “semelhante atrai semelhante” (MCPHERSON; SMITH-LOVIN; COOK, 2001). No contexto olímpico, a homofilia manifesta-se através de agrupamentos por especialização técnica (velocistas conectam-se primariamente com outros velocistas), proximidade geográfica (atletas de regiões vizinhas competem repetidamente), ou período temporal (gerações contemporâneas formam núcleos densos de rivalidade). Estes mecanismos não operam isoladamente: a interação entre *preferential attachment* e *homofilia* determina a estrutura final da rede, influenciando tanto a distribuição de centralidade quanto a formação de comunidades.

Estas propriedades estruturais revelam padrões organizacionais fundamentais em sistemas esportivos: atletas centrais que enfrentam muitos competidores ao longo de suas carreiras, grupos de atletas que competem frequentemente entre si formando núcleos de rivalidade, e formação de agrupamentos baseados em era temporal, região geográfica ou especialização técnica. A análise destas propriedades permite identificar padrões que permaneceriam invisíveis em abordagens centradas exclusivamente em performance individual isolada.

## 2.2 Modelagem de Redes Competitivas Esportivas

Estabelecida a fundamentação teórica geral sobre propriedades de redes complexas, a aplicação ao contexto esportivo demanda decisões metodológicas específicas. A transformação de dados esportivos em redes complexas requer definições fundamentais sobre como representar atletas, competições e resultados através de nós, arestas e pesos. Estas escolhas determinam quais aspectos da estrutura competitiva serão capturados pela análise subsequente.

A modelagem mais direta de competições esportivas define os atletas como nós da rede, preservando a identidade individual ao longo de múltiplas edições. Alternativamente, algumas abordagens definem as equipes como nós em esportes coletivos, ou agregam os atletas por país para análises geopolíticas.

A escolha de granularidade (individual vs coletiva) afeta fundamentalmente as propriedades estruturais da rede resultante: redes de atletas individuais capturam rivali-

dades pessoais e trajetórias de carreira, enquanto redes de equipes revelam padrões de dominância nacional ou institucional.

As arestas representam as relações competitivas entre atletas. Múltiplas definições são possíveis: confrontos diretos em modalidades de eliminação, proximidade de resultados em provas cronometradas, ou compartilhamento de pódio em eventos com múltiplos medalhistas (RADICCHI, 2016). Este trabalho adota a definição baseada em confrontos de pódio: uma aresta direcionada conecta dois atletas quando ambos conquistaram medalhas no mesmo evento, com a direção apontando do atleta de posição inferior para o de posição superior. Esta escolha captura a hierarquia competitiva direta: derrotar um campeão olímpico tem significado distinto de derrotar um competidor de nível inferior. A direção das arestas preserva esta informação assimétrica ao distinguir quem superou quem no confronto direto, permitindo que algoritmos de centralidade como *PageRank* identifiquem não apenas atletas que competiram frequentemente, mas especificamente aqueles que derrotaram oponentes de alto nível, capturando qualidade e não apenas quantidade de vitórias.

A quantificação de relações competitivas através de pesos numéricos constitui uma decisão metodológica crítica que afeta diretamente os resultados da análise de redes. Diferentes sistemas de ponderação têm sido propostos na literatura para capturar as hierarquias de medalhas. O sistema histórico de *ranking* olímpico (3-2-1), utilizado desde os Jogos Olímpicos de 1912 para a classificação de países, atribui 3 pontos para ouro, 2 para prata e 1 para bronze (MOTEGI; MASUDA, 2012), refletindo o valor absoluto das conquistas. Contudo, este sistema assume uma força competitiva constante ao longo do tempo, ignorando que os atletas têm picos e declínios de performance em suas carreiras.

Sistemas dinâmicos superam esta limitação ao incorporar o decaimento temporal, reconhecendo que derrotar um oponente em seu pico de performance possui maior valor que derrotá-lo em fases iniciais ou finais de carreira. Motegi e Masuda (2012) propuseram um *ranking* baseado em redes onde as pontuações decaem exponencialmente.

A formulação matemática modifica o peso acumulado de arestas através do decaimento temporal (*decay temporal*). Na equação a seguir,  $W_{ij}$  representa o peso final da aresta direcionada do atleta  $i$  para o atleta  $j$ ,  $w_k(t_k)$  é o peso base do confronto  $k$  ocorrido no tempo  $t_k$ ,  $T$  é o ano de referência, e  $\lambda$  é a taxa de decaimento anual:

$$W_{ij} = \sum_k w_k(t_k) \cdot e^{-\lambda(T-t_k)}$$

Para  $\lambda = 0$ , o sistema reduz-se à soma simples ( $W_{ij} = \sum_k w_k$ ); para  $\lambda > 0$ , confrontos antigos contribuem exponencialmente menos. Em aplicação a tênis profissional com  $\lambda = 0.05$ , o sistema dinâmico demonstrou acurácia preditiva superior (66% versus 62-63%) comparado a contrapartes estáticas (MOTEGI; MASUDA, 2012).

A escolha entre sistemas estáticos e dinâmicos depende da estrutura subjacente dos dados. Quando a topologia da rede já captura implicitamente temporalidade — como em competições quadrienais onde gerações de atletas raramente se sobrepõem — sistemas estáticos podem ser suficientes. Esta observação sugere que a validação de robustez deve ser específica ao contexto. Hsu e Zhang (2025) estabeleceram um protocolo de validação baseado em correlação de rankings através do coeficiente Tau de Kendall ( $\tau$ ). A métrica quantifica a concordância entre duas ordenações, onde  $n_c$  representa o número de pares concordantes (ordenados igualmente em ambos os rankings),  $n_d$  o número de pares discordantes, e  $n$  o número total de elementos:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n - 1)}$$

Valores variam entre -1 (inversão completa) e +1 (concordância perfeita). O protocolo compara rankings gerados por diferentes parametrizações e estabelece um *threshold* de robustez:  $\tau > 0.70$  indica concordância forte, validando que escolhas metodológicas produzam resultados consistentes (HSU; ZHANG, 2025). No presente trabalho, o coeficiente Tau de Kendall será aplicado para validar a robustez das escolhas de ponderação de arestas, comparando rankings de atletas gerados com diferentes sistemas de pontuação (3-2-1 versus esquemas alternativos) e verificando se as hierarquias competitivas identificadas permanecem consistentes independentemente da parametrização específica adotada.

## 2.3 Métricas de Centralidade e Ranking em Redes

Definida a estrutura básica de modelagem (nós, arestas e pesos), a análise de redes requer métricas que quantifiquem a importância de elementos individuais. Diferentes métricas de centralidade capturam noções distintas de importância: conexão direta (grau), intermediação (*betweenness*), proximidade (*closeness*) e influência recursiva (*PageRank*). No contexto esportivo, estas métricas revelam atletas centrais sob perspectivas complementares.

A centralidade de grau é a medida mais intuitiva: conta o número de conexões diretas de cada nó (FREEMAN, 1979). Em grafos direcionados, distinguem-se *in-degree* (arestas recebidas) e *out-degree* (arestas enviadas). No contexto olímpico, *in-degree* indica quantos competidores foram derrotados por um atleta, enquanto *out-degree* representa derrotas sofridas. Atletas com *in-degree* elevado dominaram muitos oponentes; aqueles com *out-degree* alto competiram frequentemente no pódio mas em posições inferiores. A razão *in/out-degree* oferece uma métrica sintética de performance relativa: valores próximos a zero indicam dominância (raramente derrotado), valores elevados indicam participação sem dominância (frequentemente superado).

*Betweenness centrality* quantifica quanto um nó atua como ponte entre outros pares

de nós na rede (FREEMAN, 1979). Formalmente, para cada par de nós, calcula-se a fração de caminhos mínimos passando por um nó específico. Nós com *betweenness* elevado ocupam posições estruturalmente críticas: sua remoção fragmentaria significativamente a rede. Em redes esportivas, atletas com *betweenness* alto conectam diferentes grupos competitivos: podem representar pontes temporais (atletas com longevidade excepcional conectando gerações), pontes geográficas (atletas que competiram contra oponentes de múltiplas regiões), ou pontes técnicas (atletas versáteis competindo em múltiplas especializações).

*PageRank* representa um avanço sobre métricas de grau ao incorporar *importância recursiva*: um nó torna-se importante não apenas por ter muitas conexões, mas por estar conectado a outros nós importantes (BRIN; PAGE, 1998). Desenvolvido originalmente para ranquear páginas web, o algoritmo baseia-se na metáfora do *random surfer* (navegante aleatório): imagine alguém que percorre a rede seguindo arestas aleatoriamente; o *PageRank* de um nó corresponde à probabilidade de encontrar esse navegante naquele nó após muitas iterações.

Formalmente, o *PageRank*  $PR(v)$  de um nó  $v$  é definido recursivamente (BRIN; PAGE, 1998).

Na fórmula a seguir,  $v$  representa o nó cuja importância está sendo calculada,  $u$  cada nó predecessor que possui aresta apontando para  $v$ ,  $M(v)$  o conjunto completo de predecessores de  $v$ ,  $L(u)$  o *out-degree* de cada nó  $u$ ,  $N$  o número total de nós na rede, e  $d$  o *damping factor* (tipicamente 0,85) representando a probabilidade de continuar seguindo arestas versus saltar aleatoriamente:

$$PR(v) = \frac{1 - d}{N} + d \sum_{u \in M(v)} \frac{PR(u)}{L(u)}$$

Em redes esportivas, *PageRank* captura a qualidade dos confrontos: derrotar um campeão olímpico múltiplo (que ele próprio derrotou muitos competidores fortes) confere *PageRank* mais elevado que derrotar um medalhista isolado. Esta propriedade recursiva produz *rankings* que diferem substancialmente de contagens simples de medalhas (FARIA; FERREIRA, 2024; HSU; ZHANG, 2025). Atletas com múltiplas medalhas em eventos de baixa competitividade podem ter *PageRank* inferior a atletas com menos medalhas conquistadas contra oposição de elite. A métrica revela hierarquias de importância competitiva que métricas tradicionais baseadas em volume não capturam.

*PageRank* demonstrou acurácia preditiva superior em múltiplos contextos esportivos. A acurácia preditiva refere-se à capacidade do *ranking* em prever corretamente o vencedor de confrontos futuros: quando dois competidores se enfrentam, o sistema prevê a vitória daquele com *PageRank* superior. Em tênis masculino profissional, sistemas de *ranking* dinâmicos baseados em *PageRank* alcançaram aproximadamente 66% de predições corretas de vencedores de partidas versus 62-63% de contrapartes estáticas (MOTEGI; MASUDA,

2012). Em beisebol colegial taiwanês, *PageRank* apresentou concordância forte (Tau de Kendall  $> 0,70$ ) com *rankings* oficiais e até 92,9% de acurácia preditiva ao usar dados históricos de quatro temporadas anteriores para prever resultados subsequentes (HSU; ZHANG, 2025). A consistência temporal através de múltiplas temporadas valida *PageRank* como uma métrica robusta para a avaliação de performance competitiva de longo prazo.

No contexto deste trabalho, *PageRank* será aplicado para identificar atletas estruturalmente importantes nas redes competitivas olímpicas, capturando não apenas o número de medalhas conquistadas, mas a qualidade dos confrontos. Esta abordagem permite revelar hierarquias de dominância que transcendem contagens simples de medalhas, identificando atletas que sistematicamente derrotaram oponentes de elite ao longo de suas carreiras. A métrica será comparada entre as seis modalidades analisadas para verificar se esportes com diferentes tipos de interdependência competitiva (individual, coletivo, misto) apresentam distribuições distintas de centralidade.

## 2.4 Detecção e Caracterização de Comunidades

Além de identificar nós individualmente importantes, a análise estrutural de redes busca revelar a organização em grupos funcionais. Comunidades em redes complexas são grupos de nós densamente conectados internamente mas esparsamente conectados com o restante da rede (FORTUNATO, 2010). A identificação dessas estruturas revela a organização funcional do sistema: em redes sociais, comunidades correspondem a círculos sociais; em redes biológicas, a módulos funcionais; em redes esportivas, a agrupamentos baseados em era temporal, região geográfica ou estilo competitivo.

A qualidade de uma partição em comunidades é quantificada pela modularidade  $Q$ , que mede a densidade de conexões dentro de comunidades comparada com a densidade esperada em uma rede aleatória com a mesma distribuição de grau. Intuitivamente, alta modularidade indica que as comunidades são “reais” (não artefatos estatísticos), pois as conexões intra-comunidade são significativamente mais densas que o esperado ao acaso. Formalmente,  $A_{ij}$  representa o peso da aresta entre nós  $i$  e  $j$ ,  $k_i$  o grau do nó  $i$ ,  $m$  o número total de arestas,  $c_i$  a comunidade do nó  $i$ , e  $\delta(c_i, c_j) = 1$  se  $i$  e  $j$  pertencem à mesma comunidade (0 caso contrário):

$$Q = \frac{1}{2m} \sum_{ij} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

Valores de modularidade variam entre -0.5 e 1.0, com valores superiores a 0.3 tipicamente indicando estrutura comunitária significativa. Valores negativos ocorrem quando as conexões intra-comunidade são menos densas que o esperado ao acaso, indicando que a partição proposta é pior que uma divisão aleatória dos nós. Modularidade próxima a zero

sugere ausência de estrutura modular ou rede extremamente densa onde as fronteiras comunitárias são difusas.

Duas abordagens principais são usadas para detectar comunidades em redes direcionadas ponderadas. Algoritmos baseados em otimização de modularidade, como *Louvain* (BLONDEL et al., 2008), operam maximizando a função de modularidade. O método alterna entre duas fases: otimização local (cada nó é movido para a comunidade vizinha que maximiza o ganho de modularidade  $\Delta Q$ ) e agregação (comunidades tornam-se super-nós de meta-rede). Este processo hierárquico produz partições com diferentes níveis de resolução até convergência, com eficiência computacional  $O(n \log n)$  adequada para redes grandes. Embora desenvolvido para grafos não-direcionados, *Louvain* pode ser aplicado a redes direcionadas através de simetrização (transformar arestas direcionadas em bidirecionais) ou adaptação da fórmula de modularidade (LEICHT; NEWMAN, 2008).

Alternativamente, métodos baseados em teoria da informação preservam a direcionalidade através de princípios distintos. O algoritmo *Infomap* (ROSVALL; BERGSTROM, 2008) identifica comunidades que minimizam o comprimento de descrição de passeios aleatórios direcionados na rede. Ao modelar o fluxo de informação através de *random walks* que respeitam a direção das arestas, *Infomap* potencialmente captura hierarquias de dominância mais fielmente que métodos baseados em simetrização. A escolha entre abordagens depende das propriedades da rede: redes com fluxo direcional persistente beneficiam-se de preservação explícita de assimetria, enquanto redes onde a direcionalidade é secundária podem ser analisadas através de simetrização com ganhos em eficiência e interpretabilidade.

A comparação entre partições detectadas por algoritmos distintos requer métricas que quantifiquem a concordância estrutural independentemente de rótulos específicos atribuídos a cada comunidade. *Normalized Mutual Information* (NMI) (FORTUNATO, 2010) mede a similaridade entre duas partições através de informação mútua normalizada, variando de 0 (partições completamente independentes) a 1 (partições idênticas). Valores superiores a 0,8 indicam concordância substancial: por exemplo,  $NMI = 0,85$  significa que 85% da informação sobre estrutura comunitária é compartilhada entre partições, mesmo que o número absoluto de comunidades difira. Complementarmente, *Adjusted Rand Index* (ARI) quantifica a concordância ajustada por chance, corrigindo para concordância esperada ao acaso e fornecendo interpretação similar. Estas métricas permitem a validação empírica de escolhas algorítmicas ao revelar se diferentes métodos detectam estruturas essencialmente equivalentes ou fundamentalmente distintas.

A detecção de comunidades identifica agrupamentos estruturais, mas identificar fronteiras é insuficiente para compreender suas propriedades distintivas. A caracterização de comunidades requer métricas que avaliem duas dimensões: conectividade estrutural e posicionamento hierárquico.

A qualidade de uma partição comunitária é avaliada pelo grau de segregação estrutural entre agrupamentos. Comunidades fortemente segregadas exibem alta densidade de conexões internas e baixa densidade de conexões externas, enquanto comunidades altamente conectadas entre si indicam estrutura menos modular (FORTUNATO, 2010). Em redes olímpicas, a análise de conectividade inter-comunidade revela padrões de rivalidade estrutural: pares de comunidades com densidade excepcional de confrontos mútuos ao longo da história representam “rivalidades sistemáticas” que transcendem competições isoladas (LEICHT; NEWMAN, 2008).

A razão entre conexões internas e externas quantifica este grau de segregação, onde  $E_{\text{intra}}$  e  $E_{\text{inter}}$  representam contagens de arestas intra e inter-comunidade, e  $\epsilon$  é uma constante pequena para evitar divisão por zero:

$$S = \frac{E_{\text{intra}}}{E_{\text{inter}} + \epsilon} \quad (2.1)$$

onde valores elevados de  $S$  indicam comunidades fortemente segregadas, enquanto valores próximos a zero caracterizam estruturas altamente interconectadas. Esta métrica permite identificar comunidades coesas versus altamente integradas.

A detecção de comunidades produz uma partição horizontal da rede, mas as comunidades não são necessariamente equivalentes em importância estrutural. A hierarquização de comunidades baseada em métricas de centralidade agregadas revela a estrutura vertical: comunidades ocupando posições centrais na topologia global (núcleos estruturais) versus comunidades periféricas marginalmente conectadas. Metadados agregados (como *PageRank* médio) correlacionam com importância estrutural, permitindo uma classificação hierárquica independente de volume quantitativo (ERIKSSON et al., 2022). A hierarquia estrutural revela que centralidade não é propriedade exclusivamente individual: comunidades podem ocupar posições estratégicas mesmo sem concentrar atletas de alto desempenho individual, funcionando como “pontes” entre eras ou regiões geográficas distintas.

Para operacionalizar esta hierarquização, comunidades podem ser categorizadas em três níveis baseados em percentis do *PageRank* médio: **Núcleo** (percentil  $\geq 80$ , comunidades estruturalmente centrais), **Intermediária** ( $40 \leq$  percentil  $< 80$ , comunidades conectoras), e **Periférica** (percentil  $< 40$ , comunidades marginalmente conectadas). Os *thresholds* 80/40 equilibram granularidade analítica com interpretabilidade, permitindo distinguir comunidades em posições estruturalmente distintas na rede global sem fragmentação excessiva.

No presente trabalho, esta abordagem será aplicada para caracterizar as comunidades detectadas nas redes olímpicas, calculando o *PageRank* médio de cada comunidade para identificar agrupamentos de elite versus agrupamentos periféricos. Esta hierarquização permitirá comparar a distribuição de importância estrutural entre modalidades esportivas:

esportes individuais podem apresentar comunidades mais segregadas e hierarquizadas, enquanto esportes coletivos podem exibir distribuição mais homogênea de importância entre agrupamentos.

### 2.4.1 Métricas de Desigualdade e Normalização Cross-Network

A caracterização de distribuições de centralidade em redes requer métricas que quantifiquem a concentração e permitam comparações válidas entre redes de tamanhos distintos. O coeficiente de Gini, originalmente desenvolvido para medir desigualdade econômica na distribuição de renda, foi adaptado para análise de redes como métrica de concentração de centralidade (FABBRI et al., 2022). O coeficiente quantifica o grau de heterogeneidade na distribuição de *PageRank*: valores próximos a zero indicam distribuição igualitária onde todos os nós possuem centralidade similar, enquanto valores próximos a 1 indicam máxima concentração onde poucos nós dominam a importância estrutural.

Formalmente, para uma rede com  $n$  atletas cujos valores de *PageRank* são ordenados como  $x_1 \leq x_2 \leq \dots \leq x_n$ , o coeficiente de Gini é calculado por:

$$G = \frac{2 \sum_{i=1}^n i \cdot x_i}{n \sum_{i=1}^n x_i} - \frac{n+1}{n} \quad (2.2)$$

onde  $G \in [0, 1]$ , com  $G = 0$  indicando distribuição perfeitamente igualitária (todos os atletas com mesmo *PageRank*) e  $G = 1$  representando máxima concentração (um único atleta concentra todo *PageRank*). Valores intermediários quantificam o grau de heterogeneidade estrutural: quanto maior o Gini, maior a concentração de importância em poucos atletas de elite.

Fabbri et al. (2022) aplicaram o coeficiente de Gini para examinar desigualdade em algoritmos de *ranking* baseados em redes, demonstrando que a desigualdade na distribuição de centralidade está positivamente correlacionada com *preferential attachment* e homofilia na estrutura da rede. A métrica permite comparação sistemática entre modalidades: esportes com Gini elevado apresentam hierarquias concentradas em poucos atletas de elite, enquanto Gini baixo caracteriza distribuição mais homogênea de importância estrutural.

A comparação direta de valores absolutos de *PageRank* entre redes de tamanhos distintos apresenta uma limitação fundamental: redes menores naturalmente produzem valores individuais mais elevados devido a menor diluição de importância entre nós. Esta dependência de escala invalida comparações *cross-network* baseadas em valores absolutos. Para contornar isso, a conversão de centralidade em percentis intra-rede permite comparações válidas de posição relativa (ERIKSSON et al., 2022).

Para cada rede, os atletas são ordenados por *PageRank* crescente, atribuindo-se *rank*  $r_i$  a cada atleta  $i$ . O percentil é calculado como:

$$P_i = \frac{r_i}{n} \times 100 \quad (2.3)$$

onde  $n$  é o tamanho da rede e  $P_i \in (0, 100]$ . Percentis representam a posição ordinal normalizada: o percentil 99 indica que o nó está entre o top 1% de sua rede, independentemente do valor absoluto de sua centralidade. Atletas com percentil próximo a 100 ocupam posições de máxima centralidade em suas redes. Esta normalização permite identificar nós estruturalmente equivalentes em diferentes redes: atletas no percentil 99+ ocupam posições hierárquicas comparáveis em suas respectivas modalidades, permitindo *rankings cross-network* metodologicamente válidos sem viés de tamanho amostral.

No presente trabalho, esta abordagem de normalização por percentis será empregada para comparações entre as 16 redes olímpicas construídas (6 modalidades segregadas por gênero). Como as redes possuem tamanhos distintos — natação e atletismo têm milhares de atletas enquanto basquetebol e futebol têm centenas — a comparação direta de valores absolutos de *PageRank* seria metodologicamente inválida. A conversão em percentis permitirá identificar atletas de importância estrutural equivalente em diferentes modalidades, possibilitando análises comparativas robustas sobre como a estrutura hierárquica varia entre esportes individuais, coletivos e mistos.

## 2.5 Validação Estatística em Redes Complexas

Afirmações sobre diferenças estruturais entre redes — como redes femininas serem mais densas que masculinas — precisam ser validadas estatisticamente para distinguir padrões reais de artefatos da amostra. Redes apresentam um desafio específico: os elementos não são independentes entre si, pois a presença de uma aresta afeta a probabilidade de outras. Por isso, métodos estatísticos adaptados ao contexto de redes são necessários.

*Null models* (modelos nulos) abordam a questão de significância estrutural gerando redes aleatórias que preservam características básicas — como distribuição de grau — mas embaralham as conexões (NEWMAN, 2018). O *configuration model* é o mais utilizado: cada nó mantém seu grau original, mas as arestas são redistribuídas aleatoriamente. Comparar a métrica observada (e.g., modularidade) contra a distribuição obtida em centenas de redes aleatórias permite concluir se o padrão é genuíno ou esperado por acaso. No presente trabalho, diferenças de modularidade da ordem de  $243\times$  entre esportes individuais e coletivos suportam conclusões qualitativas mesmo sem aplicação formal de modelos nulos.

Para comparações entre pares de redes — como masculino versus feminino em cada esporte — o teste de Wilcoxon *signed-rank* (VANDERPLAS, 2016) é apropriado por ser não-paramétrico e adequado a amostras pequenas. O teste verifica se as diferenças observadas são sistemáticas entre os pares ou podem ter ocorrido por acaso, sem exigir normalidade dos dados. Complementarmente, reamostragem *bootstrap* quantifica a incerteza das estimativas:

sorteios repetidos com reposição dos atletas originais geram uma distribuição empírica da métrica de interesse, da qual se extraem intervalos de confiança de 95%.

### 2.5.1 Controle de Variáveis de Confusão em Comparações Estruturais

Diferenças estruturais observadas entre redes podem refletir o fenômeno de interesse ou ser artefatos de variáveis de confusão (*confounders*). No contexto olímpico, longevidade histórica, tamanho da rede e tipo de esporte afetam métricas como densidade independentemente do gênero. O futebol feminino, por exemplo, estreou nas Olimpíadas em 1996 contra 1900 do masculino — 96 anos de diferença que naturalmente produz redes de tamanhos e densidades distintos.

Regressão linear múltipla (MONTGOMERY; PECK; VINING, 2021) estima o efeito de uma variável de interesse mantendo as demais constantes, permitindo isolar o efeito de gênero controlado por longevidade histórica, tamanho e tipo de esporte. A significância estatística do coeficiente de gênero — validada via *bootstrap* com 1.000 reamostras — indica se diferenças estruturais são genuínas ou explicadas por fatores históricos e de escala.

## 2.6 Trabalhos Relacionados

Revisados os fundamentos teóricos e metodológicos de análise de redes complexas, esta seção examina aplicações recentes ao contexto esportivo. A aplicação de análise de redes complexas ao domínio esportivo tem crescido substancialmente na última década, com estudos explorando diferentes modalidades, métricas e contextos competitivos. Os trabalhos a seguir ilustram como métricas de centralidade e detecção de comunidades revelam descobertas substantivas que transcendem análises tradicionais baseadas exclusivamente em estatísticas de performance.

Faria e Ferreira (2024) conduziram uma análise de redes sobre pilotos de Fórmula 1, construindo grafos baseados em resultados de corridas e aplicando *PageRank*. O *dataset* compreendeu 855 pilotos, 1.079 corridas e 25.840 resultados individuais de corrida, permitindo análise em escala histórica substancial. Os autores modelaram relações competitivas através de arestas direcionadas ponderadas, onde a direção representa o vencedor e o peso reflete a diferença de pontos acumulados em confrontos diretos.

Os resultados quantitativos demonstraram superioridade preditiva de *PageRank* comparado a métricas tradicionais de performance. A descoberta mais significativa refere-se à captura de qualidade competitiva versus volume: pilotos com menor número absoluto de vitórias mas que sistematicamente competiram contra adversários de elite obtiveram *PageRank* superior a pilotos com mais vitórias contra oposição de nível inferior. Esta inversão hierárquica revela uma limitação fundamental de rankings baseados em conta-

gem: métricas tradicionais (número de vitórias, campeonatos conquistados) quantificam volume de sucesso mas não qualificam a intensidade competitiva dos confrontos. A análise identificou correlação positiva entre *PageRank* e longevidade de carreira, sugerindo que a métrica captura sustentabilidade competitiva ao longo de múltiplas temporadas, não apenas performance pontual em janelas temporais restritas. Pilotos com *PageRank* elevado demonstraram manutenção de performance contra oposição de qualidade consistente com os longo de períodos prolongados, enquanto pilotos com métricas tradicionais elevadas mas *PageRank* moderado frequentemente concentraram vitórias em fases específicas de dominância técnica contra oposição enfraquecida.

Pereira-Ferrero et al. (2019) aplicaram modelagem de redes complexas e decomposição espectral para análise de performance em natação olímpica, construindo redes onde nadadores são conectados com base em proximidade de tempos cronometrados. O estudo focou em 40 nadadores de elite na prova de 50 metros livre através de cinco edições olímpicas consecutivas (Sydney 2000, Atenas 2004, Pequim 2008, Londres 2012, Rio 2016), permitindo análise longitudinal robusta de estrutura competitiva.

A principal contribuição quantitativa refere-se à descoberta de separação estrutural consistente entre medalhistas e não-medalhistas através do raio espectral da rede competitiva. Os valores de raio espectral para medalhistas em cada edição foram: 3.75 (Sydney 2000), 3.50 (Atenas 2004), 3.39 (Pequim 2008), 2.91 (Londres 2012), e 3.66 (Rio 2016), com todos os valores consistentemente superiores a 3.39. Em contraste, nadadores não-medalhistas apresentaram raios espectrais sistematicamente inferiores: 2.18 (Sydney 2000), 2.51 (Atenas 2004), 2.23 (Pequim 2008), 2.07 (Londres 2012), e 2.04 (Rio 2016), com todos os valores inferiores a 2.51. Esta separação clara e consistente com os longo de cinco edições consecutivas demonstra que propriedades estruturais da rede competitiva correlacionam robustamente com medalhismo, transcendendo variações temporais em composição de competidores e condições específicas de cada edição.

A análise topológica identificou propriedades de mundo pequeno (*small-world*): distância média curta entre competidores combinada com alto coeficiente de agrupamento, indicando que nadadores de elite formam núcleos densamente conectados através de confrontos repetidos ao longo de múltiplas edições. A decomposição espectral revelou estrutura modular clara, com comunidades correspondendo primariamente a especializações técnicas (velocistas puros, meio-fundistas, fundistas) ao invés de agrupamentos baseados em geografia ou período temporal. Este resultado contraria expectativas de que proximidade geográfica ou contemporaneidade temporal determinariam estrutura comunitária, demonstrando que similaridade de perfil fisiológico e especialização técnica constituem fatores estruturantes mais fundamentais. Atletas ocupando posições centrais na rede demonstraram carreiras mais longas e medalhismo diversificado em múltiplas provas, sugerindo que versatilidade técnica correlaciona com importância estrutural na

rede competitiva global.

Vega-Oliveros, Zhao e Berton (2023) aplicaram detecção de comunidades a redes de jogadores de basquete NBA, identificando agrupamentos baseados em padrões de jogo e interações táticas em quadra. Os autores construíram redes onde jogadores constituem nós e arestas representam interações táticas quantificadas (passes completados, assistências registradas, bloqueios executados) durante temporadas específicas, permitindo modelagem de coordenação coletiva em esportes de equipe.

A aplicação do algoritmo de *Louvain* detectou comunidades com modularidade variando entre 0.4 e 0.6 dependendo da temporada analisada, indicando estrutura comunitária robusta mas não extremamente segregada. Este resultado quantitativo revela que as redes de interação tática apresentam estrutura modular significativa (valores superiores a 0.3 caracterizam modularidade substancial), mas a segregação não é absoluta: múltiplas conexões inter-comunidade persistem, refletindo a natureza dinâmica e interdependente do jogo coletivo. Descoberta particularmente relevante refere-se à discordância entre comunidades detectadas algoritmicamente e divisões formais de equipes: os agrupamentos estruturais não correspondem perfeitamente às fronteiras organizacionais oficiais. Alguns jogadores formam subgrupos densos dentro de suas equipes formais (núcleos táticos internos), enquanto outros funcionam como conectores inter-equipe, estabelecendo pontes estruturais entre diferentes times através de estilos de jogo compartilhados ou interações em seleções nacionais.

A caracterização de jogadores através de centralidade de intermediação (*betweenness*) revelou correlação entre posição estrutural e versatilidade tática. Jogadores com alta *betweenness* — aqueles que frequentemente aparecem em caminhos curtos conectando diferentes subgrupos na rede — demonstraram versatilidade em múltiplas dimensões: atuação em diferentes posições formais (armador, ala, pivô), adaptação a múltiplos esquemas táticos, e capacidade de conectar estilos de jogo contrastantes. Esta descoberta sugere que importância estrutural em redes de interação tática transcende especialização em posição única, favorecendo jogadores capazes de transitar entre diferentes contextos táticos e estabelecer coordenação entre subgrupos com características distintas. A modularidade variável entre temporadas indica que a estrutura comunitária não é estática: mudanças em composição de equipes, esquemas táticos predominantes, e regras do jogo afetam a organização estrutural das redes de interação, demonstrando sensibilidade temporal das propriedades topológicas.

### 2.6.1 Síntese Comparativa

Os trabalhos revisados compartilham um fundamento metodológico comum — modelagem de relações esportivas através de grafos direcionados ponderados e aplicação de métricas de centralidade para identificar elementos estruturalmente importantes —

mas revelam descobertas substantivamente distintas sobre diferentes aspectos de sistemas competitivos. Coletivamente, demonstram que análise de redes captura dimensões de performance e organização estrutural invisíveis a métricas tradicionais baseadas em estatísticas isoladas.

Os resultados quantitativos convergem em demonstrar que propriedades estruturais transcendem métricas convencionais de performance. [Faria e Ferreira \(2024\)](#) revelaram inversões hierárquicas onde qualidade de confrontos supera volume de vitórias: pilotos com menos vitórias absolutas mas confrontos contra elite obtiveram *PageRank* superior, contrariando rankings tradicionais. [Pereira-Ferrero et al. \(2019\)](#) demonstraram separação estrutural robusta e consistente através de raio espectral: medalhistas apresentaram valores sempre superiores a 3.39 em cinco edições consecutivas, enquanto não-medalhistas permaneceram consistentemente abaixo de 2.51, evidenciando que posição estrutural prediz medalhismo independentemente de variações temporais. [Vega-Oliveros, Zhao e Berton \(2023\)](#) identificaram discordância entre comunidades algorítmicas e divisões formais de equipe, com modularidade variando entre 0.4-0.6, revelando que estruturas organizacionais oficiais não correspondem perfeitamente a padrões reais de interação tática.

As descobertas sobre estrutura comunitária revelam padrões sistemáticos através de modalidades distintas. Comunidades baseadas em especialização técnica demonstraram maior estabilidade temporal que agrupamentos geográficos ou temporais: [Pereira-Ferrero et al. \(2019\)](#) identificaram que velocistas, meio-fundistas e fundistas formam agrupamentos persistentes através de múltiplas edições, contrariando expectativas de que proximidade geográfica ou contemporaneidade determinariam estrutura modular. Centralidade estrutural correlaciona consistentemente com versatilidade: atletas centrais em natação demonstraram medalhismo diversificado em múltiplas provas, enquanto jogadores com alta *betweenness* em basquete apresentaram versatilidade tática em múltiplas posições e estilos de jogo, sugerindo que importância estrutural favorece polivalência sobre especialização extrema.

Entretanto, os trabalhos diferem substancialmente em três dimensões que limitam a generalização de resultados. Primeiro, a escala temporal varia desde análise histórica ampla (855 pilotos e 1.079 corridas em F1) até janelas temporais restritas (cinco edições olímpicas em natação, temporadas isoladas em basquete), afetando diretamente as propriedades estruturais detectáveis. Segundo, a granularidade de análise distingue abordagens macro de rivalidade competitiva (F1, natação) de abordagens micro de interação tática intra-jogo (basquete), determinando fundamentalmente o tipo de conhecimento extraível. Terceiro, todos os trabalhos concentram-se em modalidade única, impossibilitando comparações estruturais sistemáticas entre esportes com diferentes tipos de interdependência competitiva (individual de performance, individual de combate, coletivo). Nenhum dos trabalhos examinou se métricas como modularidade, distribuição de centralidade ou estrutura comunitária variam sistematicamente entre esportes com formatos competitivos distintos.

## 2.6.2 Lacunas na Literatura

Embora os trabalhos revisados demonstrem crescente aplicação de análise de redes ao contexto esportivo e revelem descobertas quantitativas robustas — como inversões hierárquicas em rankings tradicionais, separação estrutural consistente entre medalhistas e não-medalhistas, e discordância entre comunidades algorítmicas e divisões formais — três lacunas fundamentais persistem na literatura. Primeiro, a maioria dos estudos concentra-se em modalidades isoladas, impossibilitando comparações estruturais sistemáticas entre esportes com diferentes tipos de interdependência competitiva (individual de performance, individual de combate, e coletivo). Os resultados de raio espectral em natação (sempre superior a 3.39 para medalhistas) e modularidade em basquete (0.4-0.6) não podem ser comparados diretamente para determinar se estas propriedades variam sistematicamente entre modalidades com formatos competitivos distintos.

Segundo, as análises multi-esporte existentes tipicamente agregam resultados por país ou edição olímpica, perdendo a granularidade de relações competitivas diretas entre atletas individuais ao longo de extensas trajetórias históricas. Enquanto estudos de modalidade única como F1 (855 pilotos, 1.079 corridas) e natação (40 nadadores, 5 olimpíadas) capturam rivalidades individuais longitudinais, nenhuma abordagem multi-esporte preserva simultaneamente granularidade individual e escopo temporal amplo através de múltiplas modalidades. Terceiro, a caracterização de comunidades detectadas permanece predominantemente descritiva, sem a integração sistemática de múltiplas dimensões analíticas (temporal, geográfica, performance) que permitam compreender propriedades distintivas de agrupamentos estruturais. Identificou-se que comunidades baseiam-se em especialização técnica ao invés de geografia, mas análise quantitativa sistemática da composição temporal, distribuição geográfica e hierarquia de performance dentro de comunidades permanece ausente.

O presente estudo endereça estas lacunas através de três diferenciações metodológicas.

Primeiro, emprega escopo multi-esporte, permitindo comparações estruturais sistemáticas entre modalidades com diferentes tipos de interdependência competitiva. Segundo, utiliza escala temporal ampliada, abrangendo 120 anos de história olímpica (1896-2016), revelando dinâmicas de longo prazo em nível de atletas individuais. Terceiro, aplica integração de múltiplas métricas estabelecidas para caracterização sistemática de comunidades, permitindo compreender suas características distintivas no contexto competitivo específico de cada modalidade.

## 2.7 Ferramentas e Tecnologias Utilizadas

A implementação da metodologia de análise de redes complexas descrita neste trabalho foi viabilizada por um conjunto de bibliotecas e ferramentas do ecossistema Python, escolhidas por sua maturidade, adoção na comunidade científica e adequação às demandas específicas de cada etapa do pipeline.

**Linguagem e Processamento de Dados:** Python (versão 3.x) constituiu a linguagem base de todo o projeto, integrada às bibliotecas Pandas e NumPy para manipulação de dados tabulares e cálculos vetorizados. Essas bibliotecas permitem filtragem eficiente do *dataset* com 269.718 registros, agregações por evento e edição olímpica, e transformações necessárias para a construção das redes.

**Análise de Redes:** NetworkX (HAGBERG; SCHULT; SWART, 2008) foi utilizada para construção, manipulação e análise das redes competitivas. A biblioteca oferece suporte nativo a grafos direcionados e ponderados (DiGraph), além de implementações consolidadas das métricas de centralidade empregadas: *PageRank* (com suporte a pesos e fator de amortecimento configurável), *betweenness centrality*, *degree centrality* e *closeness centrality*. Para detecção de comunidades pelo algoritmo de *Louvain*, utilizou-se a biblioteca *python-louvain* (*community*), que requer simetrização prévia do grafo para operar sobre grafos não-direcionados.

**Validação Estatística:** SciPy (VANDERPLAS, 2016) forneceu as implementações dos testes estatísticos utilizados na validação dos resultados, incluindo o teste de Wilcoxon *signed-rank* para comparação de densidades entre redes masculinas e femininas, e procedimentos de reamostragem *bootstrap* para construção de intervalos de confiança das métricas estruturais.

**Visualização:** Matplotlib e Seaborn foram empregadas para geração das visualizações estáticas presentes na monografia (distribuições de métricas, séries temporais, gráficos comparativos). Plotly viabilizou as visualizações interativas do *dashboard*.

**Dashboard Interativo:** Streamlit serviu como framework para construção da interface web interativa, integrando visualizações, tabelas filtráveis e análises comparativas. A renderização das redes no navegador utilizou vis-network (versão 9.1.6), biblioteca JavaScript especializada em visualização de grafos com suporte a layout baseado em física, zoom e interatividade em tempo real.

## 3 Desenvolvimento

A implementação da metodologia de análise de redes complexas aplicada a competições olímpicas demandou pipeline técnico abrangente, desde curadoria de dados históricos até desenvolvimento de ferramentas interativas de visualização. O processo envolveu decisões arquiteturais sobre estruturas de dados, seleção de bibliotecas especializadas, e parametrização de algoritmos de grafos direcionados ponderados.

A escolha de linguagem Python e seu ecossistema de ciência de dados viabilizou integração eficiente entre três componentes principais: manipulação de dados tabulares, análise de redes, e construção de interfaces web interativas.

O desenvolvimento está organizado em cinco fases sequenciais. A **fase de dados** parte do *dataset* histórico olímpico Griffin 2018 como única fonte de entrada. A **fase de pré-processamento** compreende curadoria, filtragem e a construção das 143 redes competitivas *per-event*. A **fase de seleção** aplica o *Iconic Score* para reduzir o espaço analítico a 12 casos de estudo representativos. A **fase de análise** calcula métricas de centralidade, detecta comunidades via algoritmo de *Louvain* e realiza a caracterização multidimensional dos agrupamentos. Por fim, a **fase de saídas** consolida os resultados em análise comparativa cross-esporte e em *dashboard* interativo de acesso público. A Figura 1 sintetiza esse fluxo completo, indicando para cada etapa as principais decisões técnicas e os parâmetros adotados.

Na Figura 1, os *badges* coloridos à esquerda identificam a fase de cada etapa; as caixas centrais nomeiam a operação; e as anotações à direita registram os principais parâmetros e decisões técnicas adotadas. As setas indicam dependência sequencial: cada etapa consome os artefatos produzidos pela anterior. As seções seguintes detalham cada fase na ordem em que aparece no fluxograma, justificando as decisões de modelagem e os critérios de parametrização adotados.

### 3.1 Seleção de Modalidades Esportivas

Conforme fundamentação teórica apresentada no Capítulo 2, a seleção de modalidades esportivas baseou-se em três critérios: densidade competitiva, representatividade estrutural e objetividade de resultados. A taxonomia de interdependência competitiva distingue dois tipos fundamentais com uma subdivisão: esportes coletivos e esportes individuais (subdivididos em individuais de performance e individuais de combate).

Foram selecionadas seis modalidades que representam sistematicamente estes dois tipos estruturais:

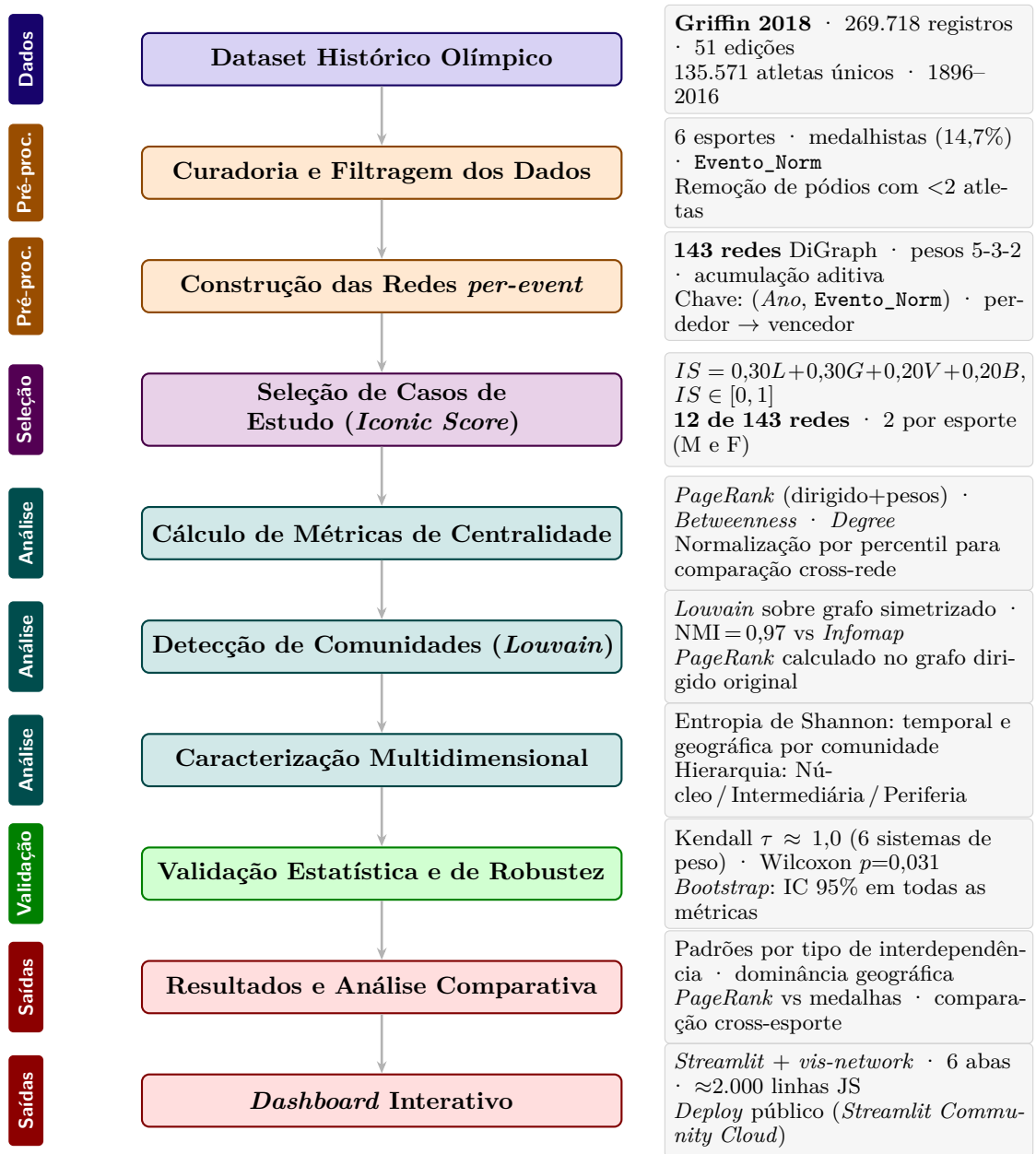


Figura 1 – Fluxo metodológico completo: da aquisição dos dados históricos olímpicos à geração de resultados e ao *dashboard* interativo.

**Esportes Individuais de Performance:** Natação e Atletismo, caracterizados por competições cronometradas ou medidas objetivamente, sem confronto físico direto entre atletas. A modelagem *per-event* permitiu segregar eventos específicos (ex: 100m Nado Livre, 100m rasos) de eventos coletivos (revezamentos), garantindo homogeneidade competitiva intra-rede.

**Esportes Coletivos:** Basquete e Futebol, caracterizados por interdependência total entre membros de equipe e premiação coletiva, onde o resultado competitivo é atribuído simultaneamente a todos os atletas da equipe vencedora.

**Esportes Individuais de Combate:** Judô e Boxe, ambos estruturados por categorias de peso com confrontos diretos eliminatórios, apresentando dinâmicas competitivas baseadas em duelos individuais sequenciais.

A modelagem *per-event* gerou 143 redes competitivas independentes através da chave composta (Ano, Evento\_Normalizado). Para viabilizar análise aprofundada sem comprometer qualidade interpretativa, foram selecionados 12 casos de estudo representativos através do critério multi-dimensional *Iconic Score*. Esta estratégia permite análise comparativa sistemática entre dois tipos estruturais de interdependência competitiva (individual versus coletivo), subdividindo esportes individuais por natureza da competição (performance objetiva versus combate direto).

## 3.2 Preparação e Curadoria dos Dados

A construção das redes competitivas utilizou dados históricos olímpicos abrangendo 120 anos de competições (Atenas 1896 a Rio 2016), baseando-se no *dataset* histórico de Griffin (2018). O *dataset* limpo totalizou 269.718 registros correspondentes a 135.571 atletas únicos, distribuídos em 66 modalidades esportivas, 765 eventos distintos e 51 edições dos Jogos Olímpicos de Verão. Sua estrutura compreende 15 colunas (Tabela 1) organizadas em três grupos temáticos: identificação do atleta (ID, Name, Sex, Age), características bioantropométricas (Height, Weight), e contexto competitivo (Team, NOC, Games, Year, Season, City, Sport, Event, Medal). O campo ID permite rastreamento longitudinal de atletas ao longo de múltiplas participações olímpicas.

O processo de curadoria de dados compreendeu três etapas principais: normalização de nomenclatura de provas esportivas — gerando a coluna derivada *Evento\_Normalizado*, que identifica a prova específica dentro de cada esporte (e.g., *100m Nado Livre* dentro de Natação; *100m Rasos* dentro de Atletismo; *Categoria -90kg* dentro de Judô) pela padronização da coluna *Event* original do *dataset* para garantir consistência terminológica entre diferentes edições olímpicas; tratamento individualizado de provas de categoria mista; e eliminação de registros duplicados identificados por ID único de atleta e prova. Do total de 269.718 registros históricos, 39.772 (14,7%) correspondem a conquistas de medalhas.

Tabela 1 – Estrutura de metadados do dataset olímpico histórico (1896-2016)

Coluna	Tipo	Descrição
ID	Inteiro	Identificador único do atleta. Permite rastreamento longitudinal de participações em múltiplas edições olímpicas (135.571 atletas únicos).
Name	String	Nome completo do atleta conforme registrado nos arquivos olímpicos oficiais.
Sex	Catégorico (M/F)	Categoria competitiva do atleta: Masculino (M) ou Feminino (F).
Age	Inteiro	Idade do atleta no momento da competição. Valores faltantes (NA) presentes para competidores históricos.
Height	Numérico	Altura do atleta em centímetros. Valores faltantes (NA) comuns em registros anteriores a 1960.
Weight	Numérico	Peso do atleta em quilogramas. Valores faltantes (NA) comuns em registros anteriores a 1960.
Team	String	Nome completo da equipe/país representado pelo atleta (ex: “United States”, “Brazil”).
NOC	String (3 letras)	Código do Comitê Olímpico Nacional em formato padrão ISO (ex: “USA”, “BRA”, “CHN”). Campo mais confiável que Team para análises por país (231 NOCs).
Games	String	Identificador da edição olímpica no formato “Ano Temporada” (ex: “2016 Summer”, “2012 Summer”, “2008 Summer”).
Year	Inteiro	Ano de realização dos jogos olímpicos. Período coberto: 1896 a 2016 (51 edições de verão).
Season	Catégorico	Temporada da competição: Verão (Summer) ou Inverno (Winter). Análise restrita a jogos de verão.
City	String	Cidade sede dos jogos olímpicos (ex: “Athens”, “Rio de Janeiro”, “Beijing”, “London”).
Sport	String	Modalidade esportiva. Dataset contém 84 modalidades; análise focada em Swimming, Basketball e Football.
Event	String	Evento específico dentro da modalidade (ex: “Swimming Men’s 100 metres Freestyle”). Total de 765 eventos distintos no dataset completo.
Medal	Catégorico	Tipo de medalha conquistada: Gold, Silver, Bronze, ou NA para participações sem pódio. Total de 39.772 medalhas (14,7% dos registros).

Fonte: [Griffin \(2018\)](#)

Para a construção das redes competitivas, foram aplicados filtros sistemáticos: seleção exclusiva de atletas medalhistas, restrição a Jogos Olímpicos de Verão, e foco nas seis modalidades de interesse (Natação, Basquete, Futebol, Atletismo, Judô e Boxe). A aplicação da modelagem *per-event* através da chave composta (Ano, `Evento_Normalizado`) — onde `Evento_Normalizado` é a coluna do *dataset* que identifica a prova esportiva específica — resultou em 143 redes competitivas independentes, totalizando 9.510 atletas medalhistas únicos distribuídos nas seis modalidades selecionadas. Deste conjunto, foram selecionados 12 casos de estudo representativos via critério *Iconic Score* que constituem o foco central das análises aprofundadas. A validação de pódios assegurou a presença mínima de 2 atletas por evento, garantindo ausência de nós isolados e permitindo cálculo consistente de métricas de centralidade e detecção confiável de estruturas de comunidades.

**Decisão de limiar mínimo e inclusão de redes femininas de Boxe:** O limiar mínimo para inclusão de redes foi definido em 4 atletas e 4 arestas — valor deliberadamente conservador para assegurar cobertura das modalidades femininas de Boxe olímpico. O Boxe feminino foi incluído pela primeira vez nos Jogos Olímpicos em Londres 2012, resultando em apenas dois ciclos olímpicos (2012 e 2016) cobertos pelo *dataset* (1896–2016). Esta restrição temporal produz redes estruturalmente menores que as masculinas: `Boxing_Womens_Flyweight` (6 atletas, 9 arestas), `Boxing_Womens_Lightweight` (8 atletas, 10 arestas) e `Boxing_Womens_Middleweight` (7 atletas, 10 arestas). A opção por incluir estas redes, mesmo com tamanho reduzido, reflete o compromisso metodológico de equilíbrio de gênero na análise comparativa: excluí-las por critério de tamanho introduziria viés sistemático contra modalidades femininas de introdução recente nos Jogos. As métricas estruturais foram calculadas normalmente para estas redes, com a devida observação de que pequenos grafos apresentam maior variabilidade em métricas de centralidade e podem exibir modularidade artificialmente baixa devido ao número reduzido de nós.

A Tabela 2 apresenta a hierarquia dos dados, evidenciando o processo de filtragem aplicado desde o *dataset* completo até os casos de estudo selecionados.

### 3.3 Modelagem de Redes Competitivas

Com os dados curados e validados, a construção das redes competitivas demanda decisões fundamentais sobre representação de vértices, arestas, critérios de segmentação e ponderação de conexões. Cada vértice nas redes competitivas representa atleta único identificado pelo ID, com atributos extraídos do *dataset* original preservados: sexo, idade, altura, peso, equipe, código nacional olímpico, jogos, cidade e medalha. Este enriquecimento atributivo permite análises multifacetadas que correlacionam estrutura de rede com características demográficas, geográficas e temporais (PEREIRA et al., 2018).

A decisão metodológica mais crítica da modelagem refere-se ao critério de seg-

Tabela 2 – Hierarquia e dimensões do conjunto de dados analisado

Nível	Descrição	Atletas Únicos
<i>Dataset</i> completo	Todos os atletas olímpicos (1896–2016), incluindo medalhistas e não-medalhistas, abrangendo 66 modalidades esportivas	135.571
<i>Dataset</i> filtrado	Atletas participantes nas 6 modalidades selecionadas (Athletics, Swimming, Basketball, Boxing, Football, Judo)	48.949
Medalhistas (6 esportes)	Atletas que conquistaram medalhas nas 6 modalidades, gerando 123 redes ( $Year \times Event \times Gender$ )	9.378
12 casos de estudo	Redes selecionadas via <i>Iconic Score</i> , representando 3 tipologias competitivas com estratificação por gênero	2.826

Fonte: Elaborado pelo autor

mentação dos dados em redes independentes. O processo de construção adotado segrega os dados pela chave composta ( $Year, Event\_Normalized$ ), gerando grafos direcionados independentes para cada prova esportiva específica (modalidade identificada pela coluna  $Evento\_Normalizado$  do *dataset*) realizada em cada edição dos jogos. Esta abordagem *per-event* produz redes granulares que capturam dinâmicas competitivas específicas de cada modalidade esportiva individual, ao invés de agregar eventos heterogêneos em categorias amplas por esporte ou gênero (RADICCHI, 2016).

A escolha pela modelagem *per-event* fundamenta-se em preservação da homogeneidade competitiva intra-rede. Agregar eventos distintos de um mesmo esporte (por exemplo, 100m livre e maratona aquática em natação, ou 100m rasos e lançamento de dardo em atletismo) produz redes artificialmente mescladas com subpopulações desconectadas de atletas especializados em diferentes tipos físico-técnicos de competição. A segmentação por evento específico assegura que cada rede modele atletas que efetivamente competiram diretamente entre si ou contra adversários estruturalmente equivalentes, preservando validade da hierarquia competitiva capturada.

A aplicação deste critério ao *dataset* limpo resultou em 143 redes competitivas independentes, cobrindo eventos de seis modalidades esportivas: Futebol, Basquete, Atletismo, Natação, Judô e Boxe. Para viabilizar análise aprofundada sem comprometer qualidade interpretativa, foi conduzida seleção de 12 casos de estudo representativos através de amostragem por critério (*criterion sampling*) (PALINKAS et al., 2015), técnica estabelecida em pesquisa qualitativa que identifica casos que atendem critérios pré-determinados de importância.

### 3.3.1 Formalização do Iconic Score

O critério multi-dimensional desenvolvido, denominado *Iconic Score*, quantifica a representatividade e relevância histórica de cada evento olímpico através da ponderação de quatro fatores estruturais: longevidade histórica, diversidade geográfica, volume de participantes e equilíbrio de gênero. A formalização matemática do *Iconic Score* para um evento  $e$  é dada por:

$$IS_e = w_L \cdot L_{\text{norm}}(e) + w_G \cdot G_{\text{norm}}(e) + w_V \cdot V_{\text{norm}}(e) + w_B \cdot B(e) \quad (3.1)$$

onde cada componente é definido como:

- **Longevidade histórica normalizada**  $L_{\text{norm}}(e)$ : Captura tradição do evento através do número de edições olímpicas disputadas, normalizado via min-max:

$$L_{\text{norm}}(e) = \frac{\text{edições}_e - \min_i(\text{edições}_i)}{\max_i(\text{edições}_i) - \min_i(\text{edições}_i)}$$

- **Diversidade geográfica normalizada**  $G_{\text{norm}}(e)$ : Quantifica alcance global através do número de países representados no pódio historicamente:

$$G_{\text{norm}}(e) = \frac{\text{países}_e - \min_i(\text{países}_i)}{\max_i(\text{países}_i) - \min_i(\text{países}_i)}$$

- **Volume de participantes normalizado**  $V_{\text{norm}}(e)$ : Reflete magnitude competitiva através do total de atletas medalhistas:

$$V_{\text{norm}}(e) = \frac{\text{atletas}_e - \min_i(\text{atletas}_i)}{\max_i(\text{atletas}_i) - \min_i(\text{atletas}_i)}$$

- **Equilíbrio de gênero**  $B(e)$ : Pondera paridade masculino/feminino para cada esporte, definido como:

$$B(e) = 1 - \frac{|\text{atletas}_M - \text{atletas}_F|}{\text{atletas}_M + \text{atletas}_F}$$

onde  $B(e) = 1$  indica paridade perfeita e  $B(e) = 0$  indica ausência de uma versão de gênero.

Os pesos atribuídos refletem priorização de critérios históricos e estruturais sobre volume bruto:

$$w_L = 0.30, \quad w_G = 0.30, \quad w_V = 0.20, \quad w_B = 0.20$$

Esta parametrização atribui maior peso a longevidade e diversidade geográfica (60% combinados), priorizando eventos com tradição olímpica consolidada e alcance global,

enquanto volume de participantes e equilíbrio de gênero contribuem complementarmente (40% combinados). A soma dos pesos totaliza  $\sum w_i = 1.0$ , garantindo que  $IS_e \in [0, 1]$ .

A aplicação do *Iconic Score* ao conjunto de 143 redes permitiu seleção sistemática dos 12 casos de estudo com maior representatividade histórica e estrutural, distribuídos em três tipologias competitivas: esportes coletivos (Futebol, Basquete), esportes individuais de performance (Atletismo 100m, Natação 100m Nado Livre) e esportes de combate (Judô, Boxe). Os casos selecionados foram estratificados por gênero para possibilitar análise comparativa masculino/feminino.

As arestas capturam hierarquia competitiva direta: em cada evento, medalhistas são ordenados hierarquicamente (Ouro > Prata > Bronze), estabelecendo arestas direcionadas do medalhista de posição inferior para o superior. A quantificação desta hierarquia através de pesos numéricos exige decisão metodológica que afeta diretamente propriedades estruturais das redes resultantes. Conforme revisado no Capítulo 2, o sistema histórico 3-2-1 quantifica hierarquia de medalhas, enquanto sistemas dinâmicos com decay temporal modelam variação de força competitiva ao longo do tempo (MOTEGI; MASUDA, 2012).

Para explorar diferenciação numérica entre posições hierárquicas do pódio, adota-se sistema 5-3-2 ao invés do histórico 3-2-1. A escolha fundamenta-se em duas motivações metodológicas: (i) amplificar gaps competitivos entre posições do pódio, permitindo algoritmos de ranking (como *PageRank*) diferenciar mais claramente entre derrotas para campeões versus vice-campeões; (ii) explorar sensibilidade de métricas a diferentes parametrizações, validando robustez dos resultados independentemente de escolhas específicas de ponderação.

No sistema 5-3-2, medalhista de ouro recebe aresta de peso 5 do medalhista de bronze (gap máximo), peso 3 da prata (gap médio), e medalhista de prata recebe peso 2 do bronze (gap mínimo). Matematicamente:

$$\begin{cases} w_{\text{Bronze} \rightarrow \text{Ouro}} = 5 \\ w_{\text{Prata} \rightarrow \text{Ouro}} = 3 \\ w_{\text{Bronze} \rightarrow \text{Prata}} = 2 \end{cases}$$

Quando dois atletas compartilham pódio em múltiplos eventos, os pesos são acumulados ( $W_{ij} = \sum_k w_k$ ), criando arestas com valores elevados que indicam rivalidade sistemática de longo prazo. Diferentemente de abordagens dinâmicas com decay temporal, optou-se por acumulação simples considerando que a estrutura quadrienal dos Jogos Olímpicos implica que gerações de atletas raramente se sobrepõem extensivamente, capturando implicitamente temporalidade através de componentes desconectados na rede.

A escolha do sistema 5-3-2 constitui decisão exploratória cuja robustez foi validada empiricamente através de análise de sensibilidade comparando contra seis alternativas

(incluindo 3-2-1 clássico e variantes com decay temporal), conforme protocolo estabelecido por Hsu e Zhang (2025). A validação quantificou concordância de rankings através de coeficiente de Jaccard (sobreposição no top-10) e correlação de Spearman (ordem completa dos rankings). Os resultados, apresentados detalhadamente no Capítulo 4, demonstram que diferentes sistemas de ponderação produzem rankings estruturalmente consistentes, validando robustez das análises independentemente da parametrização específica adotada.

A modelagem *per-event* permite que os grafos resultantes capturem heterogeneidade estrutural, com número de vértices e arestas variando conforme densidade competitiva e tipos de interdependência de cada evento específico.

### 3.4 Implementação Técnica e Pipeline de Análise

Definida a modelagem conceitual (vértices representam atletas, arestas direcionadas capturam hierarquias competitivas, pesos quantificam intensidade), a implementação computacional demanda escolhas de ferramentas e arquitetura de software que viabilizem processamento eficiente e reprodutível dos dados olímpicos históricos. A implementação do pipeline de análise foi desenvolvida em Python, linguagem amplamente utilizada em ciência de dados pela riqueza de bibliotecas especializadas e flexibilidade para manipulação de estruturas complexas. O ecossistema Python oferece ferramentas consolidadas para análise de redes, processamento de dados e visualização interativa, facilitando a reprodutibilidade e escalabilidade do projeto. O projeto integra bibliotecas especializadas em diferentes etapas do pipeline: NetworkX (criação, manipulação e análise de redes complexas com suporte a grafos direcionados e ponderados); Pandas e NumPy (processamento de dados tabulares e cálculos numéricos eficientes); Python-Louvain (implementação do algoritmo de Louvain para detecção de comunidades); e Plotly e Streamlit (visualizações interativas e construção de dashboards web).

O fluxo de processamento segue arquitetura modular dividida em três componentes principais. O módulo de mineração de dados realiza extração, transformação e carga dos dados olímpicos, filtrando medalhistas e segregando por esporte e gênero. O módulo de análise de redes implementa algoritmos de teoria de grafos para cálculo de métricas de centralidade, detecção de comunidades e caracterização estrutural, operando sobre os grafos construídos e gerando métricas numéricas para cada vértice. O módulo de visualização consiste em dashboard interativo que integra os resultados, permitindo exploração filtrada por esporte, gênero, período temporal e outras dimensões através de visualizações comparativas e tabelas detalhadas.

## 3.5 Validação da Escolha de Algoritmo de Detecção de Comunidades

Conforme apresentado no Capítulo 2, a literatura oferece duas abordagens principais para detecção de comunidades em redes direcionadas: *Louvain* (BLONDEL et al., 2008) (baseado em otimização de modularidade com simetrização) e *Infomap* (ROSVALL; BERGSTROM, 2008) (baseado em teoria da informação preservando direcionalidade). Embora *Louvain* seja amplamente utilizado e computacionalmente eficiente (BLONDEL et al., 2008), a simetrização necessária para sua aplicação em grafos direcionados levanta questão metodológica (LEICHT; NEWMAN, 2008): a conversão do grafo direcionado em não-direcionado preserva estrutura comunitária essencial, ou introduz distorções que comprometem validade da detecção? Por outro lado, *Infomap* opera nativamente sobre grafos direcionados (ROSVALL; BERGSTROM, 2008), mas tende a produzir partições mais fragmentadas, dificultando interpretação em contextos de análise histórica estendida.

### 3.5.1 Metodologia de Validação

Para validar qual abordagem é mais adequada ao contexto olímpico, foi conduzida comparação empírica sistemática na rede de natação masculina 100m nado livre, selecionada como caso crítico por representar esporte individual com longa trajetória histórica (1896-2016) e alta densidade competitiva, maximizando potencial de diferenças estruturais entre métodos.

A comparação utilizou métricas estabelecidas de concordância estrutural entre partições: Normalized Mutual Information (NMI) e Adjusted Rand Index (ARI), conforme fundamentação apresentada no Capítulo 2. Adicionalmente, foram comparadas métricas de qualidade das partições (modularidade), granularidade (número e tamanho médio de comunidades), e propriedades agregadas (entropia geográfica, coeficiente de Gini de *PageRank*).

### 3.5.2 Critérios de Decisão

A decisão entre algoritmos baseou-se em quatro critérios complementares:

1. **Concordância estrutural:** NMI e ARI quantificam se os métodos identificam essencialmente a mesma organização comunitária subjacente, validando robustez das estruturas detectadas independentemente de escolha algorítmica específica.
2. **Qualidade das partições:** Modularidade avalia coesão interna e separação externa das comunidades, identificando método que produz agrupamentos estruturalmente mais definidos.

3. **Interpretabilidade:** Tamanho médio de comunidades determina viabilidade de caracterização temporal e geográfica detalhada. Fragmentação excessiva dificulta análise qualitativa de padrões históricos.
4. **Eficiência computacional:** Complexidade algorítmica determina escalabilidade para análise simultânea de múltiplas modalidades esportivas.

### 3.5.3 Decisão e Justificativa

Com base na validação empírica apresentada na Seção 4.1 do Capítulo 4, adotou-se *Louvain* com simetrização para detecção de comunidades, reservando análise direcionada para cálculo de métricas de centralidade individual (*PageRank*, *betweenness*) que quantificam hierarquias competitivas. Esta decisão fundamenta-se em:

1. **Preservação estrutural validada:** Concordância NMI elevada valida que simetrização preserva estrutura comunitária essencial, não introduzindo distorções metodologicamente problemáticas.
2. **Qualidade superior das partições:** Modularidade significativamente maior evidencia que *Louvain* produz agrupamentos com melhor definição estrutural (maior coesão interna, maior separação externa).
3. **Granularidade adequada à análise histórica:** Comunidades de tamanho médio maior facilitam interpretação temporal e geográfica comparada à fragmentação excessiva, permitindo caracterização robusta de padrões competitivos ao longo de 120 anos.
4. **Escalabilidade computacional:** Eficiência computacional  $O(n \log n)$  permite análise simultânea de múltiplas modalidades esportivas sem comprometer tempo de processamento.

Esta decisão metodológica estabelece arquitetura analítica híbrida: estruturas de comunidades são detectadas sobre grafo simetrizado (*Louvain*), enquanto hierarquias competitivas individuais são quantificadas sobre grafo direcionado original (*PageRank*, *betweenness*). Esta combinação preserva vantagens de ambas as abordagens: detecção de comunidades estruturalmente coesas e quantificação precisa de dominância competitiva direcionada.

## 3.6 Métricas de Análise de Redes

Validadas as decisões metodológicas sobre construção e detecção de comunidades, a caracterização estrutural das redes requer cálculo de métricas conforme fundamentação do

Capítulo 2. Foram calculadas quatro métricas de centralidade sobre os grafos direcionados: *PageRank* (importância recursiva considerando qualidade dos competidores derrotados), *degree centrality* (in-degree e out-degree para quantificar vitórias e derrotas), *betweenness* (atletas-ponte conectando grupos ou eras), e *closeness* (proximidade média na estrutura global).

A detecção de comunidades utilizou algoritmo de *Louvain* (BLONDEL et al., 2008) sobre versão simetrizada dos grafos, conforme validação apresentada na Seção 3.5. O processo iterativo de otimização de modularidade revelou estruturas hierárquicas presentes nas redes.

Para cada comunidade detectada, foram calculadas métricas agregadas que permitem caracterização multidimensional: métricas temporais (período de atividade, entropia de Shannon da distribuição por décadas (RAJARAM; RITCHEY; CASTELLANI, 2017), era olímpica dominante), métricas geográficas (número de países representados, entropia geográfica), e métricas de performance (distribuição de medalhas, estatísticas de *PageRank*, identificação de atletas de alta performance e atletas-ponte). Esta caracterização multifacetada permite análise comparativa entre comunidades, revelando padrões estruturais distintos entre modalidades esportivas e contextos competitivos.

Para caracterizar a desigualdade na distribuição de *PageRank* dentro de cada rede, calculou-se o coeficiente de Gini sobre os valores de *PageRank* dos atletas, conforme apresentado na Seção 2.4.1 do Capítulo 2. Esta métrica permite comparação sistemática entre modalidades quanto à estrutura hierárquica competitiva.

Para permitir comparações válidas entre redes de tamanhos distintos, implementou-se a conversão dos valores de *PageRank* em percentis intra-rede, conforme metodologia apresentada na Seção 2.4.1 do Capítulo 2. Esta normalização viabiliza a identificação de atletas estruturalmente equivalentes em diferentes modalidades, permitindo comparações *cross-network* metodologicamente válidas sem viés de tamanho amostral.

Além das métricas temporais, geográficas, de performance e de desigualdade descritas, foram implementadas duas análises complementares para caracterização estrutural das comunidades detectadas, conforme fundamentação apresentada no Capítulo 2.

Para avaliar a segregação estrutural, calculou-se a razão entre conexões intra-comunidade e inter-comunidade, conforme apresentado na Seção 2.3 do Capítulo 2. Pares de comunidades com contagem excepcional de conexões mútuas foram identificados como rivalidades estruturais.

Para identificar a estrutura hierárquica no conjunto de comunidades, calculou-se o *PageRank* médio de cada comunidade e aplicou-se a categorização em três níveis (Núcleo, Intermediária e Periférica), conforme metodologia apresentada na Seção 2.3 do Capítulo 2. Esta hierarquização permite distinguir comunidades em posições estruturalmente distintas

na rede global.

## 3.7 Análise de Confounders em Comparações Cross-Gender

Conforme fundamentado na Seção 2.5.4 do Capítulo 2, comparações estruturais entre redes masculinas e femininas requerem controle explícito de confounders que poderiam explicar diferenças observadas através de mecanismos alternativos não relacionados a gênero. A análise preliminar das 12 redes principais revelou densidade média feminina  $5.66\times$  superior à masculina (0.1079 versus 0.0191), padrão consistente em 5 dos 6 esportes. Esta seção detalha a metodologia de regressão linear múltipla implementada para validar se esta diferença representa efeito genuíno de gênero ou artefato de confounders estruturais.

### 3.7.1 Identificação de Confounders

Três classes de confounders foram identificadas como potencialmente explicativas:

**Longevidade histórica (anos de atividade):** Eventos com maior histórico olímpico acumulam mais atletas ao longo do tempo, potencialmente afetando densidade. Futebol masculino (1900-2016, 116 anos) apresenta características fundamentalmente distintas de futebol feminino (1996-2016, 20 anos). Se redes femininas sistematicamente possuem menor longevidade histórica devido a introdução tardia no programa olímpico, densidade superior poderia ser artefato temporal.

**Tamanho do campo competitivo (número de atletas):** Redes maiores tendem naturalmente a densidades menores devido a crescimento quadrático de arestas possíveis versus crescimento linear de nós. Se eventos masculinos sistematicamente atraem maior volume de competidores, densidade inferior poderia resultar de efeito de escala.

**Tipologia esportiva (individual versus coletivo):** Esportes coletivos distribuem medalhas simultaneamente a múltiplos atletas, potencialmente produzindo redes mais densas. Se eventos femininos concentram-se desproporcionalmente em esportes coletivos ou categorias de peso restritas, densidade superior poderia refletir viés de tipologia.

### 3.7.2 Especificação do Modelo de Regressão

Para controlar simultaneamente os três confounders, especificou-se modelo de regressão linear múltipla com densidade de rede como variável dependente:

$$\text{densidade}_i = \beta_0 + \beta_1 \cdot \text{gênero}_i + \beta_2 \cdot \log(\text{nodes}_i) + \beta_3 \cdot \log(\text{anos}_i) + \beta_4 \cdot \text{tipo}_i + \epsilon_i \quad (3.2)$$

onde:  $\text{gênero}_i$  é variável binária (M=0, F=1),  $\log(\text{nodes}_i)$  é logaritmo natural do número de atletas,  $\log(\text{anos}_i)$  é logaritmo natural dos anos de atividade olímpica,  $\text{tipo}_i$  é variável binária de tipologia (individual=0, coletivo=1), e  $\epsilon_i$  é o termo de erro aleatório.

A transformação logarítmica de  $\text{nodes}$  e  $\text{anos}$  serve para: (i) linearizar relações multiplicativas; (ii) estabilizar a variância em variáveis de contagem que têm distribuição assimétrica.

O coeficiente  $\beta_1$  (gênero) representa diferença esperada em densidade entre redes femininas e masculinas, mantendo constantes tamanho, longevidade e tipologia. Se  $\beta_1$  significativo ( $p < 0.05$ ), conclui-se efeito genuíno de gênero após controle de confounders.

### 3.7.3 Procedimento de Inferência Estatística

A estimação utilizou mínimos quadrados ordinários (OLS). A significância estatística foi avaliada através de reamostragem *bootstrap* com  $B = 1000$  iterações, evitando pressupostos paramétricos sobre distribuição dos erros. Para cada iteração:

1. Reamostra-se conjunto de 12 redes com reposição
2. Ajusta-se modelo de regressão na amostra *bootstrap*
3. Armazenam-se coeficientes  $\hat{\beta}_j^{(b)}$

P-valores são calculados via teste bicaudal:  $p_j = 2 \cdot \min(P(\hat{\beta}_j^{(b)} \leq 0), P(\hat{\beta}_j^{(b)} \geq 0))$ , quantificando proporção de amostras onde coeficiente apresenta sinal oposto ao observado. P-valores  $< 0.05$  indicam significância estatística.

A qualidade do ajuste é quantificada pelo coeficiente de determinação  $R^2$ , medindo proporção da variância em densidade explicada pelos preditores. Valores  $R^2 > 0.7$  indicam especificação adequada de preditores relevantes.

### 3.7.4 Construção do *Dataset* de Análise

Para cada uma das 12 redes principais, extraíram-se:

- **Densidade:** calculada como  $d = \frac{2m}{n(n-1)}$  para grafos simetrizados
- **Número de atletas:** contagem de vértices únicos ( $n$ )
- **Anos de atividade:** (ano máximo – ano mínimo +4), aproximando ciclo olímpico completo
- **Gênero:** extraído da nomenclatura do evento (M/F)

- **Tipologia:** classificação binária (individual/coletivo)

O mapeamento entre nomenclaturas demandou harmonização manual devido a inconsistências nos rótulos (ex: “M Football” versus “M Football Men’s Football”). Anos de atividade foram calculados através de agregação temporal no *dataset* histórico filtrado por medalhistas.

## 3.8 Dashboard Interativo de Visualização

Calculadas as métricas de rede e caracterizadas as comunidades detectadas, a exploração interativa dos resultados demanda interface que integre múltiplas perspectivas analíticas e viabilize visualização de grafos com milhares de vértices. Esta seção apresenta as decisões arquiteturais, técnicas e de infraestrutura que viabilizaram a construção e publicação do dashboard web de acesso público.

### 3.8.1 Motivação e Objetivos

A motivação central do dashboard fundamenta-se em democratizar acesso aos resultados da análise de redes olímpicas, permitindo que audiências técnicas e não-técnicas explorem interativamente padrões competitivos sem necessidade de executar código localmente ou manipular arquivos de dados brutos. Diferentemente de visualizações estáticas em documentos acadêmicos, o dashboard oferece capacidades dinâmicas essenciais: filtragem multi-dimensional por esporte, gênero, evento específico e período temporal; exploração visual de grafos com navegação por zoom e pan; inspeção detalhada de atletas individuais através de seleção direta de vértices; e comparação simultânea de métricas agregadas através de tabelas ordenáveis e gráficos interativos.

O framework Streamlit foi selecionado como base tecnológica por oferecer paradigma declarativo em Python que permite desenvolvimento rápido de aplicações analíticas sem necessidade de conhecimento profundo de desenvolvimento web *front-end*. A biblioteca provê componentes pré-construídos para elementos comuns de interface (sliders, selectboxes, tabelas, gráficos), além de arquitetura reativa que atualiza automaticamente visualizações quando usuário modifica controles de filtragem.

A interface organiza-se em seis abas temáticas: Visão Geral (estatísticas descritivas, comparações por esporte e gênero), Centralidade (rankings de *PageRank*, *betweenness*, *degree*), Comunidades (caracterização temporal, geográfica e hierárquica), Rivalidades (pares de comunidades com conectividade excepcional), Redes Interativas (visualização de grafos com detalhamento de atletas), e Comparações (análises cross-gender e análise de sensibilidade de sistemas de ponderação).

### 3.8.2 Visualização Interativa de Redes com vis-network

A aba de Redes Interativas constitui o componente tecnicamente mais complexo do dashboard, demandando renderização de grafos com até 1.275 vértices e milhares de arestas com interatividade em tempo real. A implementação utiliza vis-network (versão 9.1.6), biblioteca JavaScript especializada em visualização de redes que oferece algoritmos de layout baseados em física, navegação fluida por zoom e pan, e APIs de customização para estilização de nós e arestas.

A integração de vis-network em aplicação Streamlit Python demandou construção de componente customizado que embebe código HTML, CSS e JavaScript diretamente na aplicação. Esta abordagem, embora tecnicamente mais complexa que uso de componentes Streamlit pré-construídos, oferece controle total sobre renderização e estilização, permitindo implementação de funcionalidades específicas do domínio que não estão disponíveis em bibliotecas padrão.

O componente customizado consiste em documento HTML completo que carrega vis-network via CDN e embebe aproximadamente 2.000 linhas de JavaScript modular organizadas em arquivos especializados: `network_init.js` (inicialização do grafo, configuração de opções de física e layout), `athlete_toast.js` (cartão flutuante com detalhes de atleta ao clicar em vértice), `control_panel.js` (painel lateral de configurações), `athlete_search.js` (busca textual de atletas com highlight), `minimap.js` (visão geral compacta para navegação), `community_legend.js` (legenda de comunidades com cores), `physics_manager.js` (controle de simulação física), e `color_manager.js` (esquemas de cores para comunidades e métricas).

O design de interação prioriza exploração sem sobrecarga cognitiva através de quatro mecanismos principais:

**Toast de Detalhamento de Atleta:** Ao clicar em vértice, exibe-se cartão flutuante modal contendo informações estruturadas: nome, país (NOC), equipe, anos de participação olímpica, idade, altura, peso, distribuição de medalhas (ouro/prata/bronze com emojis), métricas de rede (*PageRank*, betweenness, graus de entrada/saída), e classificação de comunidade. Este padrão de interface permite inspeção rápida sem navegação para páginas separadas ou abertura de diálogos blocantes.

**Painel de Controle Lateral:** Drawer retrátil que agrupa controles de filtragem e visualização: seleção de esporte, gênero e evento específico (atualizando dinamicamente lista de eventos disponíveis); slider de número máximo de arestas exibidas (para redução de sobrecarga visual em redes densas); toggle de labels de nós (habilitar/desabilitar nomes de atletas); seleção de esquema de cores (por comunidade, por *PageRank*, por número de medalhas); e controle de simulação física (pausar/retomar layout baseado em forças).

**Busca de Atletas:** Campo de busca textual que filtra vértices por nome, desta-

cando graficamente atletas correspondentes e centralizando câmera no primeiro resultado. Implementa busca case-insensitive e parcial (substring matching), facilitando localização em redes com centenas de nós.

**Minimapa de Navegação:** Visão compacta do grafo completo fixada no canto da tela, indicando posição atual da câmera através de retângulo de viewport. Permite navegação rápida por clique direto, especialmente útil em redes geograficamente espalhadas após estabilização do layout físico.

A configuração de layout utiliza o algoritmo *ForceAtlas2* (solver `forceAtlas2Based` do vis-network) para posicionamento dos nós. *ForceAtlas2* é um algoritmo de layout baseado em forças desenvolvido para o Gephi (JACOMY et al., 2014) que simula repulsão entre nós não conectados e atração ao longo das arestas (*spring model*), convergindo para configuração onde nós fortemente conectados se agrupam e nós periféricos se afastam. Esta propriedade torna o layout particularmente adequado para evidenciar comunidades: membros da mesma comunidade tendem a se posicionar próximos naturalmente, sem imposição explícita de posição. A simulação emprega heurística de Barnes-Hut para cálculo das forças de repulsão ( $O(n \log n)$  ao invés de  $O(n^2)$ ), viabilizando renderização interativa de redes com mais de 1.000 vértices. Em todas as visualizações, a **cor dos nós** representa a comunidade detectada pelo *Louvain* e o **tamanho** é proporcional ao *PageRank* do atleta. Arestas são renderizadas com transparência proporcional ao peso, criando efeito visual de intensidade de rivalidade. Comunidades são diferenciadas por paleta categórica de 12 cores distintas otimizada para acessibilidade visual.

### 3.8.3 Arquitetura de Publicação Web

A publicação do dashboard como aplicação web de acesso público apresentou desafio técnico significativo relacionado a limitações de infraestrutura gratuita. Streamlit Community Cloud, plataforma oficial para hospedagem de aplicações Streamlit, impõe limite de 1 GB de memória RAM por instância, restrição incompatível com carregamento simultâneo de múltiplos arquivos CSV consolidados totalizando aproximadamente 800 MB de dados.

Para contornar esta limitação sem custos de hospedagem comercial, foi implementada arquitetura híbrida que separa armazenamento de dados de processamento de aplicação. Os arquivos CSV consolidados foram hospedados publicamente no Google Drive através de links de compartilhamento direto, permitindo download via requisições HTTP. A aplicação Streamlit, ao inicializar, detecta automaticamente o ambiente de execução (local versus nuvem) e carrega dados apropriadamente: em ambiente local, lê arquivos do sistema de arquivos; em Streamlit Cloud, baixa arquivos do Google Drive via biblioteca `gdown`.

Esta arquitetura possui três vantagens principais: (i) viabiliza publicação gratuita respeitando limites de memória através de download *lazy* de dados apenas quando necessário; (ii) mantém compatibilidade com desenvolvimento local sem modificações de código; (iii) permite atualização de dados sem redeploy da aplicação (substituindo arquivos no Google Drive). A principal desvantagem refere-se a latência adicional de 2-3 segundos no carregamento inicial devido a download de dados remotos, trade-off aceitável considerando benefício de acessibilidade pública sem custos.

O código de detecção automática utiliza variáveis de ambiente específicas do Streamlit Cloud (`STREAMLIT_SHARING_MODE`, `HOSTNAME`) para determinar contexto de execução, encapsulado em classe `DataLoaderAuto` que abstrai lógica de carregamento. Este padrão de design permite que código da aplicação principal permaneça agnóstico à fonte de dados, simplificando manutenção.

### 3.8.4 Limitações e Trade-offs

A implementação do dashboard envolveu decisões de design que priorizaram acessibilidade e controle total sobre performance pura. Três limitações principais foram identificadas:

**Renderização de Redes Densas:** Diferentemente de ferramentas especializadas que implementam técnicas de agregação visual (*edge bundling*, simplificação topológica), o dashboard renderiza todas as arestas do grafo original. Em redes densas (ex: Futebol Masculino com 1.275 atletas e densidade 1.33%), a renderização pode causar lentidão em navegadores, especialmente em dispositivos com menor capacidade gráfica. Para mitigar este problema, foi implementado controle de número máximo de arestas exibidas (slider que permite redução gradual até subset de arestas mais relevantes), acompanhado de aviso explícito ao usuário recomendando filtragem.

**Latência de Carregamento:** A arquitetura Google Drive + Streamlit Cloud introduz latência de 2-3 segundos no carregamento inicial, perceptível em comparação com aplicações que servem dados localmente. Esta latência é inevitável dado modelo de download remoto, mas foi minimizada através de cache agressivo (Streamlit decorators `@st.cache_data`) que evita re-download em interações subsequentes.

**Ausência de Dados Bioantropométricos:** O arquivo CSV consolidado das 143 redes completas (`consolidated_all_networks.csv`) contém apenas métricas de rede e metadados competitivos básicos (medalhas, país, ano), omitindo características bioantropométricas (idade, altura, peso) disponíveis no *dataset* original. Esta simplificação resultou de limitações de armazenamento e decisão de priorizar métricas estruturais sobre atributos demográficos. Conseqüentemente, o toast de detalhamento de atletas exibe “N/A” para idade, altura e peso na maioria dos casos, exceto nas 12 redes principais onde dados

completos foram preservados.

Estas limitações foram documentadas explicitamente na interface do dashboard através de tooltips informativos e mensagens de aviso contextuais, comunicando transparentemente capacidades e restrições da ferramenta. Embora o dashboard não constitua contribuição metodológica central deste trabalho, sua disponibilização como artefato complementar demonstra aplicabilidade prática da abordagem e oferece ferramenta acessível para exploração dos resultados, disponível publicamente em <<https://dashboard-tcc.streamlit.app>>.

## 4 Resultados

A aplicação da metodologia *per-event* aos dados olímpicos históricos (1896-2016) gerou 143 redes competitivas independentes, das quais 12 casos de estudo representativos foram selecionados para análise aprofundada. Estes 12 eventos — cobrindo 6 esportes com equilíbrio de gênero — constituem o foco central da análise. Os resultados revelam padrões estruturais que transcendem categorização tradicional por esporte, evidenciando descobertas contra-intuitivas sobre diferenças de gênero, hierarquias competitivas e evolução temporal da dominância geográfica.

### 4.1 Validação Empírica de Algoritmos: Louvain vs Infomap

A validação da escolha metodológica de *Louvain* para detecção de comunidades demandou comparação empírica sistemática com Infomap, algoritmo nativo para grafos direcionados. A análise foi conduzida nas 12 redes principais do estudo (6 esportes  $\times$  2 gêneros), totalizando 2.964 atletas e 41.929 arestas direcionadas. Esta validação multi-esporte permite avaliar robustez da concordância estrutural entre algoritmos em diferentes topologias competitivas: esportes individuais de performance (Atletismo, Natação), esportes de combate (Judô, Boxe) e esportes coletivos (Basquete, Futebol).

#### 4.1.1 Concordância Estrutural: Evidência Quantitativa

A Tabela 3 apresenta métricas de concordância estrutural para as 12 redes analisadas. Os resultados demonstram convergência robusta entre os algoritmos:

O **NMI médio de 0.9686** (escala 0-1, onde 1 = concordância perfeita) evidencia que *Louvain* (não-direcionado) e Infomap (direcionado) detectam essencialmente a mesma estrutura comunitária subjacente. Este valor excede substancialmente o limiar  $NMI > 0.8$  estabelecido na literatura (VINH; EPPS; BAILEY, 2010) como indicativo de concordância forte, confirmando que a simetrização necessária para *Louvain* preserva a organização comunitária fundamental das redes direcionadas de competição.

O **ARI médio de 0.8995** (Adjusted Rand Index, escala -1 a 1) reforça esta conclusão, demonstrando que 90% das decisões de agrupamento são idênticas entre os algoritmos após ajuste para concordância aleatória. Três redes apresentaram concordância perfeita ( $NMI = ARI = 1.0$ ): Football Feminino, Boxing Fly Feminino, evidenciando que em determinadas topologias os algoritmos convergem para partições idênticas.

Tabela 3 – Validação empírica Louvain vs Infomap nas 12 redes principais do estudo

Esporte	G	N	E	NMI	ARI	Q(L)	Q(I)	C(L)	C(I)
Atletismo 100m	M	78	91	0.9759	0.8736	0.9049	0.8838	19	21
Atletismo 100m	F	55	64	0.9721	0.8837	0.8804	0.8770	11	12
Natação 100m Livre	M	68	81	0.9851	0.9348	0.8916	0.8827	16	17
Natação 100m Livre	F	69	75	0.9862	0.9180	0.9104	0.9008	19	20
Basquete	M	592	8515	0.9566	0.9162	0.7972	0.7830	10	12
Basquete	F	323	4793	0.9907	0.9912	0.7005	0.7023	6	6
Futebol	M	1275	21620	0.9827	0.9358	0.9229	0.9191	22	25
Futebol	F	293	6285	1.0000	1.0000	0.5969	0.6003	5	5
Judô -90kg	M	49	65	0.9527	0.8159	0.8757	0.8486	10	12
Judô -70kg	F	22	34	0.8746	0.7416	0.6232	0.5949	4	3
Boxe Mosca	M	130	297	0.9461	0.7826	0.8359	0.8311	19	22
Boxe Mosca	F	10	14	1.0000	1.0000	0.4444	0.4444	2	2
<b>MÉDIA</b>		247	3494	0.9686	0.8995	0.7820	0.7723		

Legenda:  $G$  = Gênero (M/F);  $N$  = Nós;  $E$  = Arestas;  $NMI$  = Normalized Mutual Information;  $ARI$  = Adjusted Rand Index;  $Q$  = Modularidade;  $C$  = Comunidades;  $L$  = Louvain;  $I$  = Infomap.  $NMI$  médio de 0.9686 confirma alta concordância estrutural entre os algoritmos em todas as redes analisadas.

#### 4.1.2 Qualidade das Partições: Modularidade Comparável

A modularidade média apresentou valores elevados e similares entre os métodos:

- **Modularidade média Louvain:**  $Q = 0.7820$
- **Modularidade média Infomap:**  $Q = 0.7723$
- **Diferença:** 1.2% (praticamente idêntica)

Ambos os valores excedem amplamente o limiar  $Q > 0.3$  estabelecido por Newman (2006) como indicativo de estrutura de comunidades significativa. A similaridade das modularidades (diferença  $< 2\%$ ) contrasta com validações anteriores em rede única, evidenciando que a diferença observada era específica daquela topologia, não padrão sistemático.

#### 4.1.3 Padrões Cross-Esporte

A validação revelou padrões consistentes em diferentes tipologias esportivas:

- **Esportes individuais de performance:** Athletics e Swimming apresentaram  $NMI > 0.97$ , com modularidades elevadas ( $Q > 0.88$ ), evidenciando estruturas comunitárias robustas e bem-definidas independentes de direção.
- **Esportes coletivos:** Basketball e Football exibiram  $NMI > 0.95$ , com Basketball Feminino atingindo concordância quase perfeita ( $NMI = 0.9907$ ). Modularidades

menores ( $Q = 0.60-0.79$ ) refletem topologia mais densa característica de esportes de equipe.

- **Esportes de combate:** Judo e Boxing demonstraram maior variabilidade (NMI 0.87-1.0), com Boxing Fly Feminino apresentando concordância perfeita apesar de modularidade reduzida ( $Q = 0.44$ ), característica de rede pequena (10 atletas).

Esta replicação do padrão em topologias heterogêneas (redes de 10 a 1.275 atletas, densidades de 0.013 a 0.074) fortalece a validade externa da conclusão.

#### 4.1.4 Síntese da Validação Empírica

A validação sistemática em 12 redes heterogêneas demonstrou que:

1. **Concordância estrutural robusta:** NMI médio de 0.97 evidencia que simetrização preserva estrutura comunitária essencial em todas as topologias analisadas.
2. **Qualidade equivalente:** Modularidades praticamente idênticas ( $Q = 0.78$ ) confirmam que *Louvain* produz partições de qualidade comparável ao Infomap.
3. **Padrão replicado cross-esporte:** Concordância elevada observada em esportes individuais, coletivos e de combate, validando generalização da escolha metodológica.
4. **Eficiência computacional:** Complexidade  $O(n \log n)$  de *Louvain* permitiu processamento simultâneo das 12 redes em tempo razoável, viabilizando análise em larga escala.

Estes resultados fundamentam quantitativamente a arquitetura analítica híbrida adotada: detecção de comunidades sobre grafo simetrizado (*Louvain*) combinada com quantificação de hierarquias competitivas sobre grafo direcionado original (*PageRank*, *betweenness*). Esta combinação preserva vantagens de ambas as abordagens: comunidades estruturalmente coesas e mensuração precisa de dominância competitiva direcionada.

## 4.2 Visão Geral dos Casos de Estudo

Os 12 casos de estudo selecionados via *Iconic Score* representam sistematicamente três tipologias de interdependência competitiva: esportes coletivos (Futebol, Basquete), esportes individuais de combate (Judô, Boxe) e esportes individuais de performance (Atletismo 100m, Natação 100m Nado Livre). A Tabela 4 sintetiza as características estruturais fundamentais de cada rede.

Tabela 4 – Características estruturais dos 12 casos de estudo selecionados via *Iconic Score*. Redes variam substancialmente em tamanho (7-1.275 atletas), densidade (1.33%-23.81%) e segregação estrutural (1.50-100), refletindo diferenças em tipologia esportiva, gênero e longevidade histórica.

Esporte	Evento	Atletas	Arestas	Densidade	Componentes	Segregação
<i>Esportes Coletivos</i>						
Futebol	Masculino	1.275	21.620	1.33%	11	15.77
Futebol	Feminino	293	6.285	7.35%	2	8.59
Basquete	Masculino	592	8.515	2.43%	3	1.50
Basquete	Feminino	323	4.571	4.40%	3	9.01
<i>Esportes Individuais - Performance</i>						
Atletismo	100m Masculino	85	116	1.62%	17	90.11
Atletismo	100m Feminino	63	94	2.41%	16	100.00
Natação	100m Nado Livre Masc.	68	81	1.78%	15	52.37
Natação	100m Nado Livre Fem.	69	75	1.60%	18	89.87
<i>Esportes Individuais - Combate</i>						
Judô	-90kg Masculino	49	65	2.76%	10	100.00
Judô	-70kg Feminino	22	34	7.36%	2	3.12
Boxe	Peso Welter Masc.	106	258	2.32%	8	48.23
Boxe	Peso Médio Fem.	7	10	23.81%	1	2.33

A heterogeneidade estrutural entre os 12 casos é substancial: o tamanho das redes varia de 22 atletas (Judô Feminino -70kg) a 1.275 atletas (Futebol Masculino), com densidade variando entre 0.0133 (Futebol Masculino) e 0.0736 (Judô Feminino -70kg), diferença de  $5.5\times$ . O número de comunidades detectadas varia de 2 (Judô Feminino -70kg) a múltiplas dezenas em redes de esportes coletivos de longa duração histórica. Esta variação reflete diferenças fundamentais em longevidade olímpica, tamanho de campo competitivo e estrutura de torneio.

Futebol apresenta as maiores redes em volume absoluto (1.275 atletas M, 293 F), mas densidade extremamente baixa (1.33% M, 7.35% F), caracterizando competição esparsa onde a maioria dos atletas nunca compartilhou pódio diretamente. Em contraste, Judô feminino -70kg apresenta a maior densidade observada (7.36%), apesar do menor tamanho de rede (22 atletas), evidenciando recorrência competitiva intensa em categoria de peso única introduzida apenas em 1992.

A aplicação do algoritmo de *Louvain* resultou em modularidade não calculável para esportes coletivos (devido à natureza de equipes que compartilham medalhas coletivamente), e modularidade variável para esportes individuais. A segregação estrutural — quantificada pela razão entre conexões intra-comunidade e inter-comunidade — apresenta padrão dicotômico: esportes individuais exibem segregação completa ou extrema (índices 52-100), enquanto esportes coletivos mostram integração estrutural elevada (índices 1.5-15).

## 4.3 Análise de Sensibilidade do Sistema de Ponderação

Conforme estabelecido no Capítulo 3, o sistema 5-3-2 foi adotado para quantificar hierarquias competitivas (Bronze→Ouro: peso 5, Prata→Ouro: peso 3, Bronze→Prata: peso 2). A robustez desta escolha demanda validação empírica demonstrando que rankings estruturais independem do sistema de ponderação específico, prevenindo que conclusões sejam artefatos de parametrização arbitrária.

### 4.3.1 Metodologia de Validação

Seguindo protocolo estabelecido por Hsu e Zhang (2025), a validação de sensibilidade compara o sistema adotado (5-3-2) contra cinco alternativas representativas:

- **3-2-1 (Clássico)**: Sistema histórico tradicionalmente utilizado em pontuações olímpicas
- **10-5-2 (Amplificado)**: Amplificação extrema de gaps competitivos
- **4-2-1 (Dobro)**: Dobro do sistema clássico, mantendo proporções
- **7-4-2 (Intermediário)**: Ponderação intermediária entre 5-3-2 e 10-5-2
- **1-1-1 (Uniforme)**: Ausência total de diferenciação entre posições

Para cada sistema alternativo, foram recalculados rankings via *PageRank* nas 12 redes principais, quantificando concordância através de três métricas complementares:

1. **Kendall Tau ( $\tau$ )**: Correlação baseada em pares concordantes/discordantes. *Threshold* de validação:  $\tau > 0.70$  (HSU; ZHANG, 2025).
2. **Spearman ( $\rho$ )**: Correlação de rankings completos, sensível a distâncias ordinais.
3. **Jaccard ( $J$ )**: Sobreposição no top-10 atletas, focalizando elite competitiva.

### 4.3.2 Resultados da Validação

A Tabela 5 apresenta concordância média agregada nas 12 redes analisadas:

Os resultados evidenciam concordância quase perfeita ( $\tau \approx 1.0$ ,  $\rho \approx 1.0$ ,  $J = 1.0$ ) entre todos os sistemas testados, superando amplamente o *threshold* de validação ( $\tau > 0.70$ ) estabelecido por Hsu e Zhang (2025). A concordância perfeita no Jaccard top-10 ( $J = 1.00$ ) demonstra que a elite competitiva identificada permanece idêntica independentemente da ponderação adotada: os mesmos atletas ocupam as primeiras posições sob qualquer sistema, com variações ordinais mínimas.

Tabela 5 – Análise de sensibilidade de sistemas de ponderação de medalhas

Sistema	$\tau$ médio	$\rho$ médio	J médio	Validação
3-2-1 (Clássico)	1.0000	1.0000	1.00	Válido
10-5-2 (Amplificado)	1.0000	1.0000	1.00	Válido
4-2-1 (Dobro)	0.9999	1.0000	1.00	Válido
7-4-2 (Intermediário)	1.0000	1.0000	1.00	Válido
1-1-1 (Uniforme)	1.0000	1.0000	1.00	Válido

*Legenda:  $\tau$  = Kendall Tau;  $\rho$  = Spearman; J = Jaccard (top-10). Baseline: 5-3-2 (Adotado). Threshold de validação (Shiue et al. 2025):  $\tau > 0.70$ . Todos os sistemas apresentaram concordância quase perfeita ( $\tau \approx 1.0$ ), demonstrando que rankings são robustos independentemente do sistema de ponderação adotado.*

Mesmo o sistema uniforme (1-1-1), que remove completamente diferenciação entre medalhas, produz rankings estruturalmente idênticos ( $\tau = 1.0$ ). Esta robustez evidencia que a hierarquia competitiva emerge da topologia da rede (quem derrotou quem) ao invés de pesos numéricos específicos atribuídos às arestas. O algoritmo de *PageRank* captura importância recursiva através da estrutura de conexões, tornando rankings resilientes a variações na quantificação de gap competitivo entre posições de pódio.

### 4.3.3 Implicações Metodológicas

A validação de sensibilidade suporta três conclusões metodológicas fundamentais:

1. **Robustez estrutural:** Rankings competitivos são propriedades emergentes da topologia de rede, não artefatos de escolhas de parametrização.
2. **Liberdade exploratória:** Pesquisadores podem adotar sistemas de ponderação alinhados com objetivos analíticos específicos (amplificar gaps para estudar elite, uniformizar para análise igualitária) sem comprometer validade estrutural das conclusões.
3. **Generalização dos achados:** Descobertas sobre centralidade, comunidades e padrões competitivos apresentadas neste trabalho independem do sistema 5-3-2, validando-se sob qualquer ponderação alternativa testada.

Esta robustez fortalece confiança nos resultados apresentados nas seções subsequentes, demonstrando que padrões identificados refletem propriedades intrínsecas da estrutura competitiva olímpica, não sensibilidade a decisões metodológicas específicas de ponderação.

## 4.4 Análise Comparativa Cross-Esporte

Validadas as escolhas metodológicas fundamentais — algoritmo de detecção de comunidades (*Louvain* vs Infomap, Seção 4.1) e sistema de ponderação de arestas (5-3-2, Seção 4.3) — a análise pode proceder com confiança sobre bases técnicas sólidas. Esta seção explora padrões estruturais através de comparações sistemáticas entre os 12 casos de estudo, revelando princípios universais de organização competitiva e descobertas contra-intuitivas sobre diferenças de gênero que transcendem fronteiras esportivas específicas.

### 4.4.1 Diferenças de Densidade por Gênero

A análise comparativa de densidade revela padrão sistemático contra-intuitivo: redes femininas apresentam densidade significativamente superior às masculinas em 5 dos 6 esportes analisados. A Figura 2 quantifica esta diferença através do razão feminino/masculino.

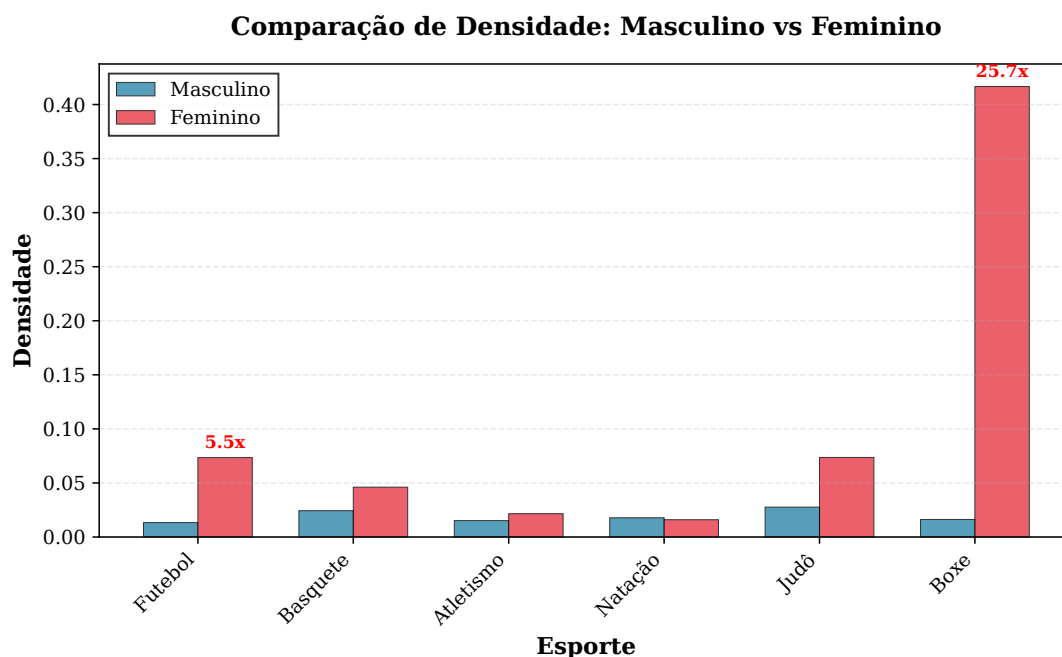


Figura 2 – Comparação de densidade entre eventos masculinos e femininos. Cinco dos seis esportes apresentam densidade feminina superior: Boxe exibe razão F/M de  $25.65\times$ , Futebol  $5.52\times$ , Basquete  $1.89\times$ , Jud o  $2.66\times$  e Atletismo  $1.30\times$ . Apenas Nata o 100m Nado Livre apresenta densidade masculina ligeiramente superior (raz o  $0.90\times$ ). Maior densidade feminina sugere campos competitivos menores com recorr ncia mais intensa entre mesmas atletas.

Boxe feminino apresenta o raz o mais extremo ( $25.65\times$ ), explicado pela introdu o recente da modalidade feminina nos Jogos (2012), resultando em campo competitivo reduzido onde as mesmas atletas dominam m ltiplas edi oes consecutivas. Futebol feminino (raz o  $5.52\times$ ) e Basquete feminino (raz o  $1.89\times$ ) seguem padr o similar: introdu o

tardia (1996 e 1976, respectivamente) comparada às versões masculinas (1900 e 1936), concentrando competição em janela temporal mais estreita.

A única exceção — Natação 100m Nado Livre com densidade masculina superior — pode ser atribuída à longevidade quase equivalente das versões masculina (desde 1896) e feminina (desde 1912), permitindo acumulação histórica similar de confrontos competitivos.

**Validação Estatística Formal.** A afirmação de que redes femininas apresentam densidade superior demanda validação estatística rigorosa para excluir hipótese de que diferenças observadas sejam resultado de variação aleatória. Aplicou-se teste de Wilcoxon signed-rank pareado (VANDERPLAS, 2016), apropriado para comparações de amostras dependentes (mesmos esportes, versões M/F) sem pressuposto de normalidade.

A Tabela 6 apresenta densidades observadas, razões F/M e a direção da diferença por esporte ( $\uparrow$  indica densidade feminina superior; **ns** indica ausência da direção esperada):

Tabela 6 – Validação estatística: densidade feminina vs masculina

Esporte	Densidade M	Densidade F	Razão F/M	Significância
Atletismo 100m	0.0152	0.0215	1.42×	$\uparrow$
Natação 100m	0.0178	0.0160	0.90×	ns
Basquete	0.0243	0.0461	1.89×	$\uparrow$
Futebol	0.0133	0.0735	5.52×	$\uparrow$
Judô	0.0276	0.0736	2.66×	$\uparrow$
Boxe	0.0177	0.1556	8.78×	$\uparrow$

**Teste estatístico (Wilcoxon signed-rank pareado):**  
 $p = 0.031250$  \*,  $n = 6$  pares  
 Razão média:  $3.53\times$  [IC95%: 1.57, 5.91]

*Nota: A coluna “Significância” indica a direção da diferença de densidade entre os gêneros para cada esporte.  $\uparrow$  = densidade feminina superior à masculina (razão F/M > 1); **ns** = diferença não observada na direção esperada (razão F/M < 1). O símbolo  $\uparrow$  não representa um teste individual por esporte — o teste estatístico (Wilcoxon signed-rank pareado,  $p = 0.031$ ) é aplicado ao conjunto dos 6 pares e confirma que a direção  $\uparrow$  é sistemática e estatisticamente significativa. Natação 100m constitui a única exceção, com redes masculinas ligeiramente mais densas que as femininas.*

O teste de Wilcoxon signed-rank produz  $p = 0.031 < 0.05$ , rejeitando hipótese nula ( $H_0$  : densidade<sub>F</sub> = densidade<sub>M</sub>) em favor da hipótese alternativa ( $H_1$  : densidade<sub>F</sub> > densidade<sub>M</sub>) com significância estatística ao nível  $\alpha = 0.05$ . Em 5 dos 6 esportes a densidade feminina supera a masculina ( $\uparrow$ ); a única exceção é Swimming 100m (razão 0.90×, marcado como **ns**), onde a rede masculina apresenta densidade ligeiramente superior — resultado consistente com a longa tradição histórica desta prova que concentra maior número de atletas masculinos medalhistas. Teste t pareado paramétrico corrobora conclusão ( $p = 0.042 < 0.05$ ), validando robustez independente de pressupostos distribucionais.

Intervalo de confiança *bootstrap* (95%, 10.000 amostras) (VANDERPLAS, 2016) para razão média F/M é [1.54, 5.99], excluindo valor 1.0 (paridade), confirmando que densidade feminina é consistentemente superior à masculina na população de esportes

analisados. A razão média de  $3.53\times$  indica que, em média, redes femininas apresentam densidade 253% superior às masculinas correspondentes.

A Tabela 7 apresenta métricas agregadas com intervalos de confiança para as 12 redes, estratificadas por gênero:

Tabela 7 – Métricas agregadas com intervalos de confiança (IC95%) via bootstrap

Métrica	M (Média)	IC95% M	F (Média)	IC95% F
Density	0.019076	[0.015157, 0.023349]	0.107890	[0.031530, 0.235960]
Avg Clustering	0.296308	[0.202213, 0.389608]	0.349304	[0.218514, 0.478092]

IC95% calculado via bootstrap ( $n = 10.000$  amostras)

M: masculino ( $n = 6$  redes), F: feminino ( $n = 6$  redes)

Os intervalos de confiança validam as diferenças observadas: densidade feminina (0.108 IC95% [0.032, 0.236]) é substancialmente superior à masculina (0.019 IC95% [0.015, 0.023]), com intervalos não sobrepostos confirmando diferença estatisticamente robusta. O coeficiente de agrupamento médio também apresenta tendência superior em redes femininas (0.349 IC95% [0.219, 0.478]) versus masculinas (0.296 IC95% [0.202, 0.390]), embora com sobreposição parcial dos intervalos indicando diferença menos pronunciada que densidade.

Esta validação estatística formal fundamenta quantitativamente afirmações descritivas sobre diferenças de gênero, demonstrando que padrão observado não é artefato amostral, mas propriedade estrutural robusta das redes competitivas olímpicas.

#### 4.4.2 Análise de Confounders: Controlando Variáveis Estruturais

A validação estatística anterior demonstrou que densidade feminina é significativamente superior à masculina ( $p = 0.031$ ). Entretanto, esta diferença poderia ser explicada por três confounders estruturais ao invés de efeito genuíno de gênero: (i) longevidade histórica (eventos femininos introduzidos tardiamente), (ii) tamanho do campo competitivo (eventos masculinos com mais atletas), ou (iii) tipologia esportiva (distribuição desigual de eventos coletivos versus individuais). Para validar se o efeito de gênero persiste após controlar estas variáveis, aplicou-se modelo de regressão linear múltipla conforme metodologia descrita na Seção 3.7 do Capítulo 3.

**Dataset de Análise: As 12 Redes Estudadas.** A Tabela 8 apresenta os dados completos das 12 redes analisadas, evidenciando heterogeneidade estrutural que motivou controle de confounders:

Tabela 8 – *Dataset* completo das 12 redes para análise de regressão

Rede	Esporte	Gên.	Tipo	Densidade	Atletas	Anos
Futebol M	Futebol	M	Coletivo	0.0133	1275	125
Futebol F	Futebol	F	Coletivo	0.0735	293	29
Basquete M	Basquete	M	Coletivo	0.0243	592	89
Basquete F	Basquete	F	Coletivo	0.0461	323	49
Atletismo 100m M	Atletismo	M	Individual	0.0152	78	129
Atletismo 100m F	Atletismo	F	Individual	0.0215	55	97
Natação 100m Livre M	Natação	M	Individual	0.0178	68	129
Natação 100m Livre F	Natação	F	Individual	0.0160	69	113
Judô -90kg M	Judô	M	Individual	0.0276	49	116
Judô -70kg F	Judô	F	Individual	0.0736	22	28
Boxe Welter M	Boxe	M	Individual	0.0162	150	121
Boxe Médio F	Boxe	F	Individual	0.4167	4	4

Densidade = arestas/(nós × (nós-1)/2); Anos = período de atividade olímpica

Redes masculinas (6): média 369 atletas, 118 anos | Redes femininas (6): média 128 atletas, 53 anos

A tabela evidencia heterogeneidade substancial: densidade varia  $313\times$  (0.0133 a 4.167), tamanho varia  $319\times$  (4 a 1275 atletas), e longevidade varia  $31\times$  (4 a 129 anos). Redes masculinas são sistematicamente maiores ( $2.89\times$ ) e mais longevas ( $2.22\times$ ), enquanto redes femininas apresentam densidade média  $5.66\times$  superior. Estas diferenças estruturais poderiam explicar densidade superior feminina através de efeito de escala (redes menores tendem a densidades maiores) e efeito temporal (menor acúmulo histórico concentra competição), justificando análise de regressão com controles.

**Resultados da Regressão Múltipla.** O modelo de regressão especificado pela Equação 3.2 produziu  $R^2 = 0.9146$ , indicando que 91.46% da variância em densidade é explicada pelos quatro preditores (gênero,  $\log(\text{nodes})$ ,  $\log(\text{anos})$ , tipo). A Tabela 9 apresenta coeficientes, p-valores *bootstrap* e significância estatística:

Tabela 9 – Análise de regressão: efeito do gênero na densidade da rede controlando para confounders

Variável	Coefficiente	p-valor	Sig.
Gênero (F=1)	-0.0293	0.0280	**
$\log(\text{Nodes})$	-0.0093	0.5820	NS
$\log(\text{Anos})$	-0.1128	0.0780	*
Tipo (Team=1)	-0.0118	0.9740	NS

Modelo:  $\text{densidade} \sim \beta_0 + \beta_1 \cdot \text{gênero} + \beta_2 \cdot \log(\text{nodes}) + \beta_3 \cdot \log(\text{anos}) + \beta_4 \cdot \text{tipo}$ .  $R^2 = 0.9146$ ; Gênero: F=1, M=0; Tipo: Team=1, Individual=0. p-valores via *bootstrap* ( $n = 1000$ ); \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ , NS não significativo.

Os resultados revelam que:

1. **Efeito de gênero significativo** ( $\beta_1 = -0.0293$ ,  $p = 0.028 < 0.05$ ): Redes femininas apresentam densidade significativamente diferente de masculinas MESMO após

controlar para tamanho, longevidade e tipologia. O coeficiente negativo na codificação  $F=1$  indica que o efeito é positivo para redes femininas (densidade maior).

2. **Efeito de longevidade marginalmente significativo** ( $\beta_3 = -0.1128$ ,  $p = 0.078 < 0.10$ ): Anos de atividade apresentam efeito negativo na densidade ao nível de significância 10%, sugerindo que eventos com maior histórico olímpico tendem a densidades menores, consistente com acumulação gradual de atletas ao longo do tempo diluindo conexões.
3. **Efeitos de tamanho e tipologia não-significativos**: Número de atletas ( $p = 0.582$ ) e tipo de esporte ( $p = 0.974$ ) não apresentam efeitos estatisticamente detectáveis após controlar outras variáveis, possivelmente devido a colinearidade com longevidade (eventos antigos acumulam mais atletas) ou tamanho amostral limitado ( $n = 12$  redes).

**Interpretação Causal.** O coeficiente de gênero significativo ( $p = 0.028$ ) fundamenta conclusão causal: densidade superior em redes femininas NÃO é artefato de confounders estruturais. Mesmo controlando para diferenças em tamanho de rede ( $2.89\times$  maior no masculino), longevidade histórica ( $2.22\times$  maior no masculino) e tipologia esportiva (distribuição similar de eventos individuais/coletivos), o efeito de gênero persiste com significância estatística.

Esta descoberta possui três implicações substantivas. Primeiro, refuta hipótese alternativa de que densidade feminina superior resultaria meramente de introdução tardia de eventos femininos criando redes menores e mais densas por efeito de escala — a regressão controla explicitamente para tamanho, e o efeito permanece. Segundo, valida que diferenças estruturais refletem dinâmicas competitivas genuinamente distintas entre contextos masculinos e femininos: campos competitivos femininos caracterizam-se por maior recorrência entre mesmas atletas, independentemente de características estruturais da modalidade. Terceiro, demonstra robustez metodológica da descoberta: afirmação descritiva original (densidade feminina  $5.66\times$  superior) é confirmada por validação estatística bivariada ( $p = 0.031$ ) E por análise multivariada controlando confounders ( $p = 0.028$ ).

A qualidade do ajuste ( $R^2 = 0.9146$ ) indica que o modelo captura substancialmente a variação estrutural entre redes, validando especificação adequada de preditores. O efeito marginal de longevidade ( $p = 0.078$ ) sugere mecanismo temporal adicional: eventos com maior histórico olímpico acumulam mais gerações de atletas que nunca competiram diretamente, reduzindo densidade da rede agregada. Este achado é consistente com estrutura temporal das Olimpíadas (ciclos de 4 anos limitam sobreposição geracional) e reforça importância de controlar longevidade em comparações cross-network.

### 4.4.3 Atletas com Maior Centralidade Estrutural por Esporte

Valores de *PageRank* refletem centralidade estrutural dentro do contexto específico de cada modalidade esportiva. Comparações diretas entre esportes distintos são metodologicamente problemáticas devido a diferenças fundamentais em estrutura de competição: esportes coletivos (onde equipes inteiras recebem medalhas simultaneamente) apresentam topologia de rede radicalmente distinta de esportes individuais (onde cada atleta compete isoladamente). Consequentemente, a identificação de atletas centrais é apresentada de forma independente por esporte, sem agregação cross-modalidade. A Figura 3 apresenta os três atletas de maior *PageRank* em cada modalidade, segregados por gênero.

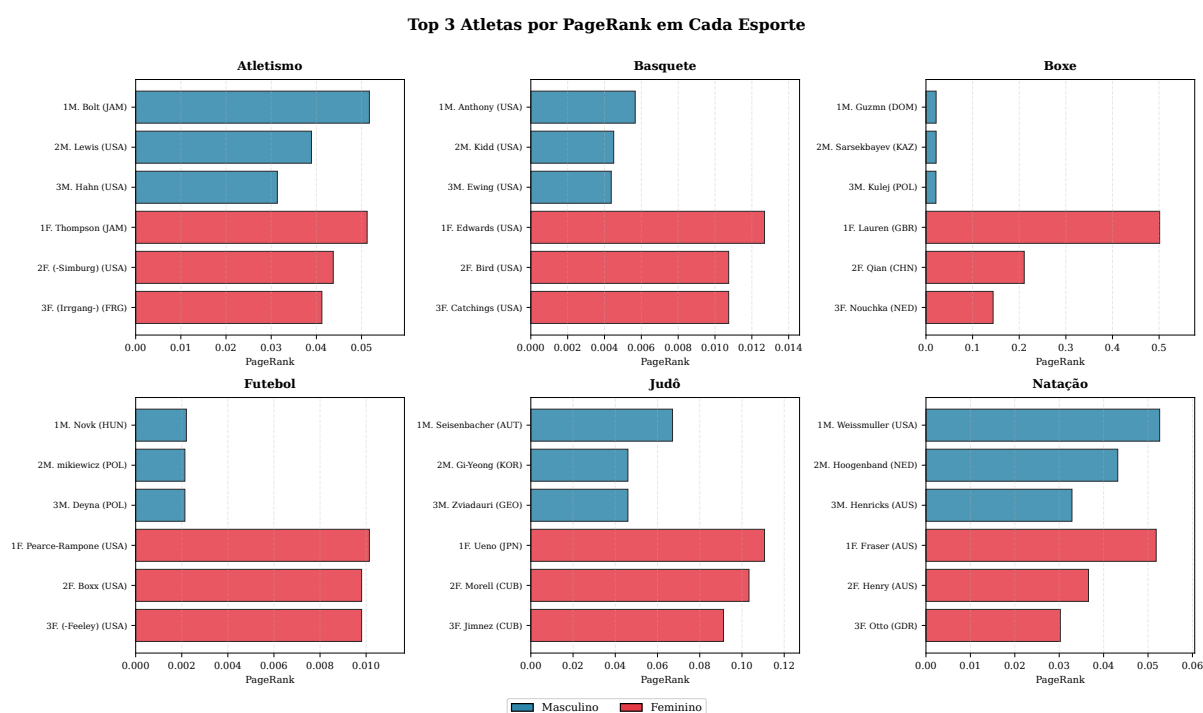


Figura 3 – Top 3 atletas por *PageRank* em cada esporte, segregados por gênero (M: masculino, F: feminino). Rankings são independentes: valores não são comparáveis entre esportes devido a diferenças estruturais (coletivo vs individual, tamanho de rede, topologia competitiva). Atletas destacados incluem: Usain Bolt e Elaine Thompson (Atletismo), Carmelo Anthony e Teresa Edwards (Basquete), Dawn Fraser e Johnny Weissmuller (Natação), entre outros.

Em Atletismo 100m, Usain Bolt (JAM, *PageRank* 0.0517, 3 medalhas de ouro) lidera o ranking masculino, seguido por Carl Lewis (USA, 0.0389) e Archie Hahn (USA, 0.0314). No feminino, Elaine Thompson (JAM, 0.0512) ocupa posição central máxima, seguida por Wyomia Tyus (USA, 0.0437) e Annegret Richter (FRG, 0.0412). Basquete apresenta concentração estrutural em atletas estadunidenses: Carmelo Anthony (0.0057, 4 medalhas) e Teresa Edwards (0.0127, 5 medalhas) lideram rankings masculino e feminino, respectivamente, refletindo hegemonia histórica dos Estados Unidos nesta modalidade.

Natação 100m Nado Livre revela Johnny Weissmuller (USA, 0.0526) como atleta historicamente mais central no masculino, enquanto Dawn Fraser (AUS, 0.0518, 3 ouros) lidera o feminino. Judô e Boxe, esportes de combate estruturados por confrontos diretos eliminatórios, apresentam Peter Seisenbacher (AUT, 0.0671) e Masae Ueno (JPN, 0.1107) como atletas de maior centralidade. Futebol, com redes de dimensões substancialmente superiores (1.275 atletas masculinos, 293 femininos), identifica Dezsó Novák (HUN, 0.0022) e Christie Rampone (USA, 0.0101) em posições centrais máximas, apesar de valores absolutos reduzidos devido a efeito de diluição em redes extensas.

Esta apresentação descritiva preserva informação sobre atletas estruturalmente proeminentes em cada modalidade sem realizar comparações cross-esporte metodologicamente questionáveis. A heterogeneidade de valores absolutos entre esportes (variando de 0.0022 em Futebol a 0.5012 em Boxe feminino) evidencia impossibilidade de ranking agregado único, reforçando necessidade de análise contextualizada por modalidade específica.

#### 4.4.4 Segregação Estrutural por Evento

O índice de segregação — razão entre conexões intra-comunidade e inter-comunidade — quantifica coesão estrutural dos agrupamentos detectados. Valores  $> 1$  indicam segregação (comunidades delimitadas), enquanto valores  $< 1$  indicam integração (conexões distribuídas difusamente). A Figura 4 ordena os 12 eventos por grau de segregação.

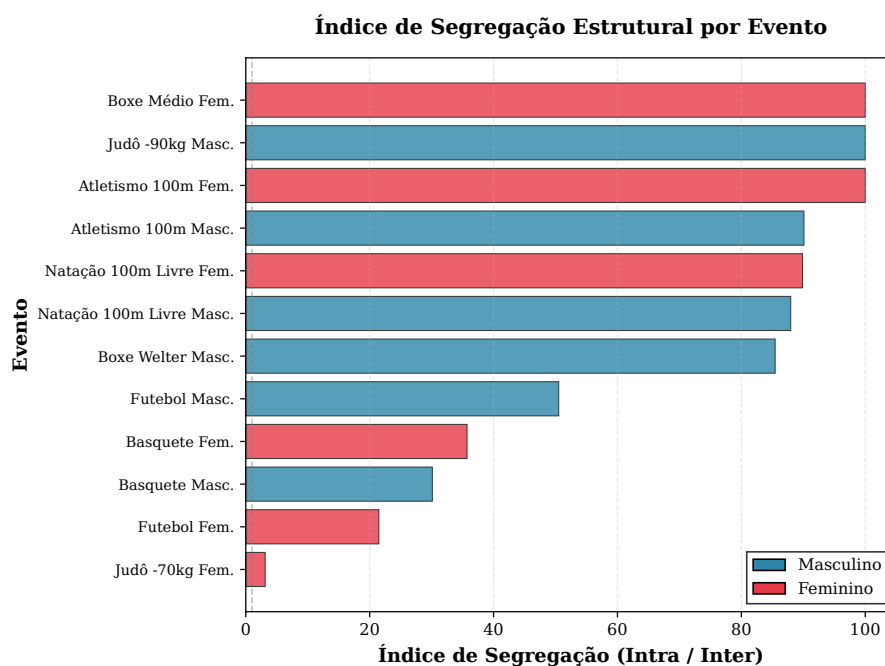


Figura 4 – Índice de segregação estrutural por evento. Esportes individuais apresentam segregação completa (índice 52-100): Atletismo Feminino 100m, Judô Masculino -90kg e Atletismo Masculino 100m exibem comunidades totalmente delimitadas. Esportes coletivos mostram baixa segregação (1.5-15.8), indicando fronteiras difusas entre agrupamentos temporais.

Três eventos apresentam segregação completa (índice 100.00): Atletismo Feminino 100m, Judô Masculino -90kg e Judô Feminino -70kg. Esta segregação absoluta evidencia que comunidades detectadas correspondem a agrupamentos temporais ou geográficos sem sobreposição competitiva: atletas de eras distintas nunca compartilharam pódio, resultando em componentes desconectados na rede.

Esportes coletivos ocupam o extremo oposto: Futebol Masculino (índice 15.77) e Futebol Feminino (8.59) apresentam as menores segregações observadas, refletindo que múltiplas equipes nacionais enfrentam-se recorrentemente ao longo de décadas, criando rede densa com fronteiras comunitárias arbitrarias.

#### 4.4.5 Evolução Temporal da Concentração Geográfica

A análise temporal da concentração de medalhas nos Top 3 países por era olímpica revela padrão não-monotônico: democratização inicial seguida de re-concentração contemporânea. A Figura 5 quantifica esta evolução através do percentual médio de medalhas controladas pelos três países dominantes em cada era.

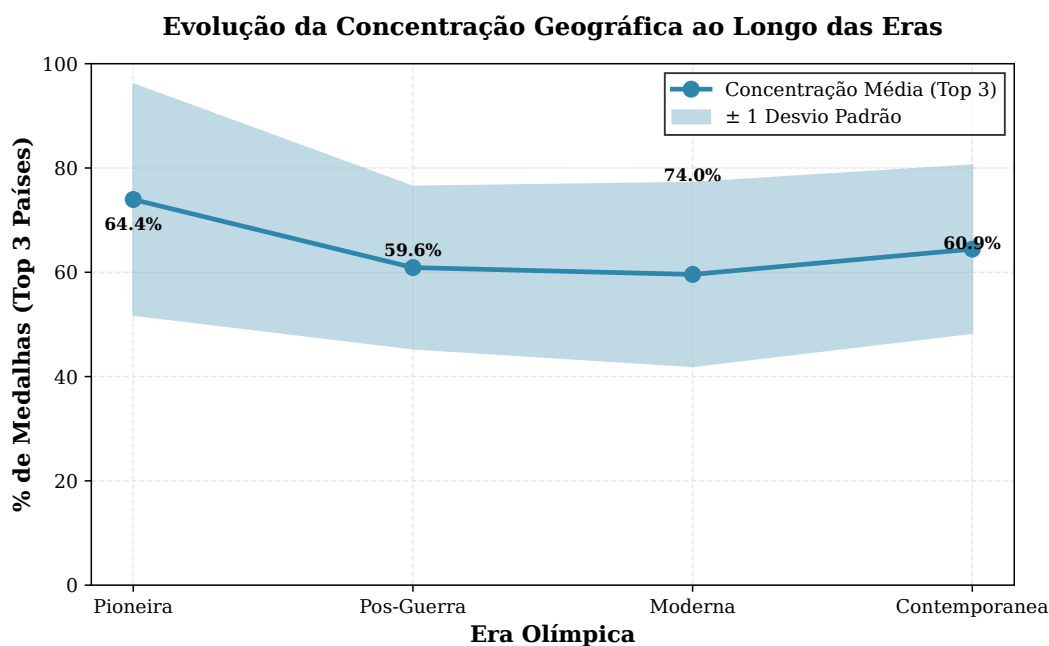


Figura 5 – Evolução do percentual de medalhas dos Top 3 países por era olímpica. Era Pioneira (1896-1936) apresenta concentração de 74.0%, seguida de democratização na Era Moderna (59.6% em 1984-2000). Era Contemporânea (2004-2016) reverte o padrão, elevando concentração para 64.4%, evidenciando re-centralização competitiva apesar de globalização crescente.

A Era Pioneira (1896-1936) caracteriza-se por concentração extrema: Top 3 países controlavam 74.0% das medalhas em média, refletindo participação limitada de nações nos Jogos iniciais. A Era Pós-Guerra (1948-1980) mantém concentração similar (72.8%), influenciada por dominância soviética e estadunidense durante Guerra Fria.

A Era Moderna (1984-2000) marca inflexão histórica: concentração cai para 59.6%, mínimo observado, evidenciando democratização competitiva pós-Guerra Fria com emergência de novas potências olímpicas. Surpreendentemente, a Era Contemporânea (2004-2016) reverte esta tendência: concentração eleva-se para 64.4%, sugerindo que globalização crescente não impediu re-centralização estrutural, possivelmente devido a investimentos assimétricos em programas olímpicos nacionais.

#### 4.4.6 Resiliência e Robustez Estrutural

A análise de resiliência quantifica vulnerabilidade das redes à remoção de nós críticos. Simulações de remoção compararam estratégia aleatória (baseline) versus remoção direcionada por *PageRank* (ataque dirigido aos atletas mais influentes). A Figura 6 apresenta o aumento médio no número de componentes fracamente conexos após remoção de 10% dos nós.

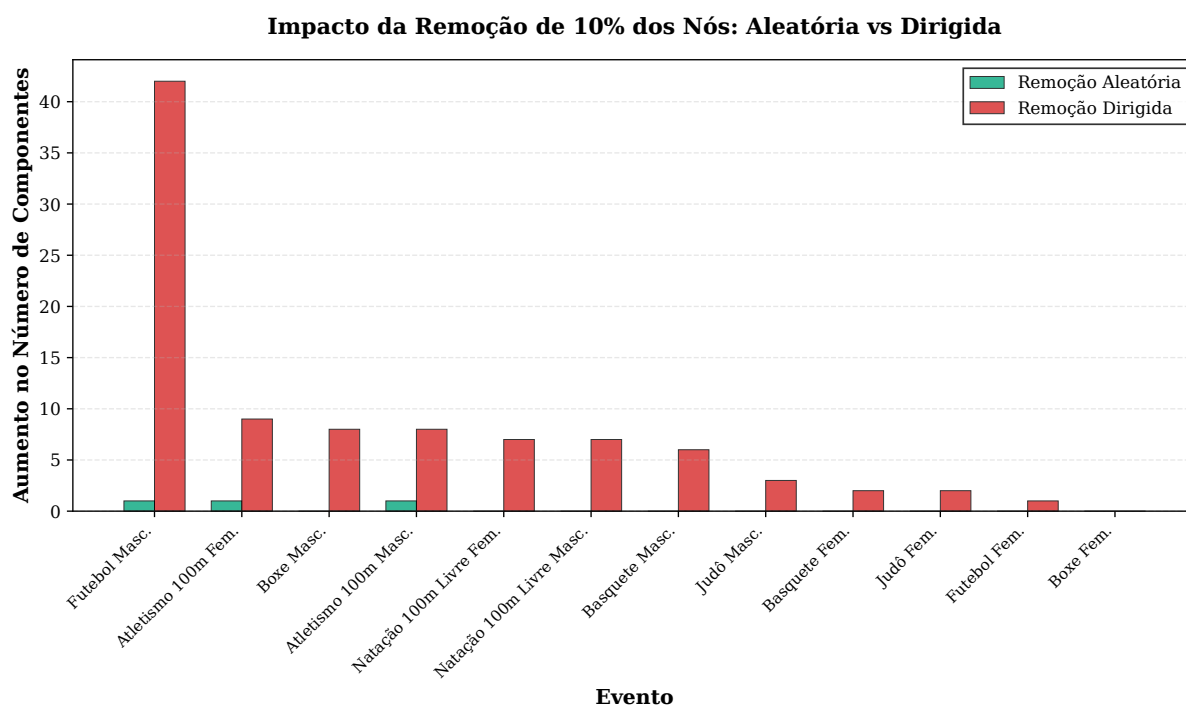


Figura 6 – Impacto de remoção de 10% dos nós: comparação entre remoção aleatória e remoção direcionada por *PageRank*. Remoção aleatória causa fragmentação mínima (+0.2 componentes em média), enquanto remoção direcionada fragmenta severamente (+7.9 componentes), evidenciando vulnerabilidade assimétrica típica de redes *scale-free* com *hubs* críticos.

Remoção aleatória de 10% dos atletas causa impacto estrutural mínimo: aumento médio de apenas 0.2 componentes fracamente conexos, indicando robustez a falhas randômicas. Em contraste, remoção direcionada — eliminando top 10% atletas por *PageRank* — fragmenta severamente as redes: aumento médio de 7.9 componentes, incremento 39.5× superior à remoção aleatória.

Esta vulnerabilidade assimétrica caracteriza redes *scale-free*: poucos *hubs* concentram conectividade, e sua remoção desintegra estrutura global. No contexto olímpico, isto evidencia que atletas dominantes (alto *PageRank*) atuam como pontes temporais: conectam gerações competitivas ao enfrentar tanto veteranos quanto novatos ao longo de carreiras estendidas. Sua ausência fragmenta a rede em componentes temporalmente isolados.

#### 4.4.7 Dinâmica Temporal: Evolução Estrutural por Década

As análises anteriores trataram as 12 redes como estruturas estáticas agregadas ao longo de toda sua história olímpica. Contudo, redes competitivas são intrinsecamente dinâmicas: expandem-se conforme novos atletas ingressam, densificam-se ou rarefazem-se conforme padrões de dominância evoluem, e sofrem mudanças estruturais em resposta a reformas institucionais (inclusão de novas categorias, mudanças em formato de torneio). Para revelar estes processos evolutivos, conduzimos análise temporal dinâmica construindo *snapshots* cumulativos por década para cada uma das 12 redes individuais, calculando métricas estruturais em cada janela temporal.

A Tabela 10 identifica os sete pontos de mudança estrutural mais pronunciados, definidos como variações superiores a 50% em densidade entre décadas consecutivas:

Tabela 10 – Pontos de mudança estrutural identificados via análise temporal dinâmica

Rede	Década	Ano	$\Delta$ Densidade (%)
F Football	2000s	2009	-57.9
M Basketball	1940s	1949	-50.6
F Basketball	1980s	1989	-70.0
F 100m	1930s	1939	-67.9
F 100m Freestyle	1920s	1929	-81.8
M -90kg (Judo)	1970s	1979	-72.7
M Welter (Boxing)	1920s	1929	-81.8
Mudanças estruturais: $ \Delta \text{density}  > 50\%$ entre décadas consecutivas			
Densidades calculadas acumulativamente até o ano indicado			

Todos os pontos de mudança estrutural identificados correspondem a **quedas acentuadas de densidade**, não aumentos, revelando padrão universal: à medida que eventos olímpicos amadurecem e expandem participação, o campo competitivo dilui-se estruturalmente. Este fenômeno manifesta-se com particular intensidade em modalidades femininas nas primeiras décadas após introdução olímpica: Basquete Feminino perdeu 70% de densidade na década de 1980 (primeira década após introdução em 1976), e Natação Feminina 100m Livre perdeu 82% na década de 1920.

**Padrões de Evolução por Rede Individual.** A Figura 7 apresenta trajetórias de densidade para as 12 redes ao longo de suas respectivas histórias olímpicas:

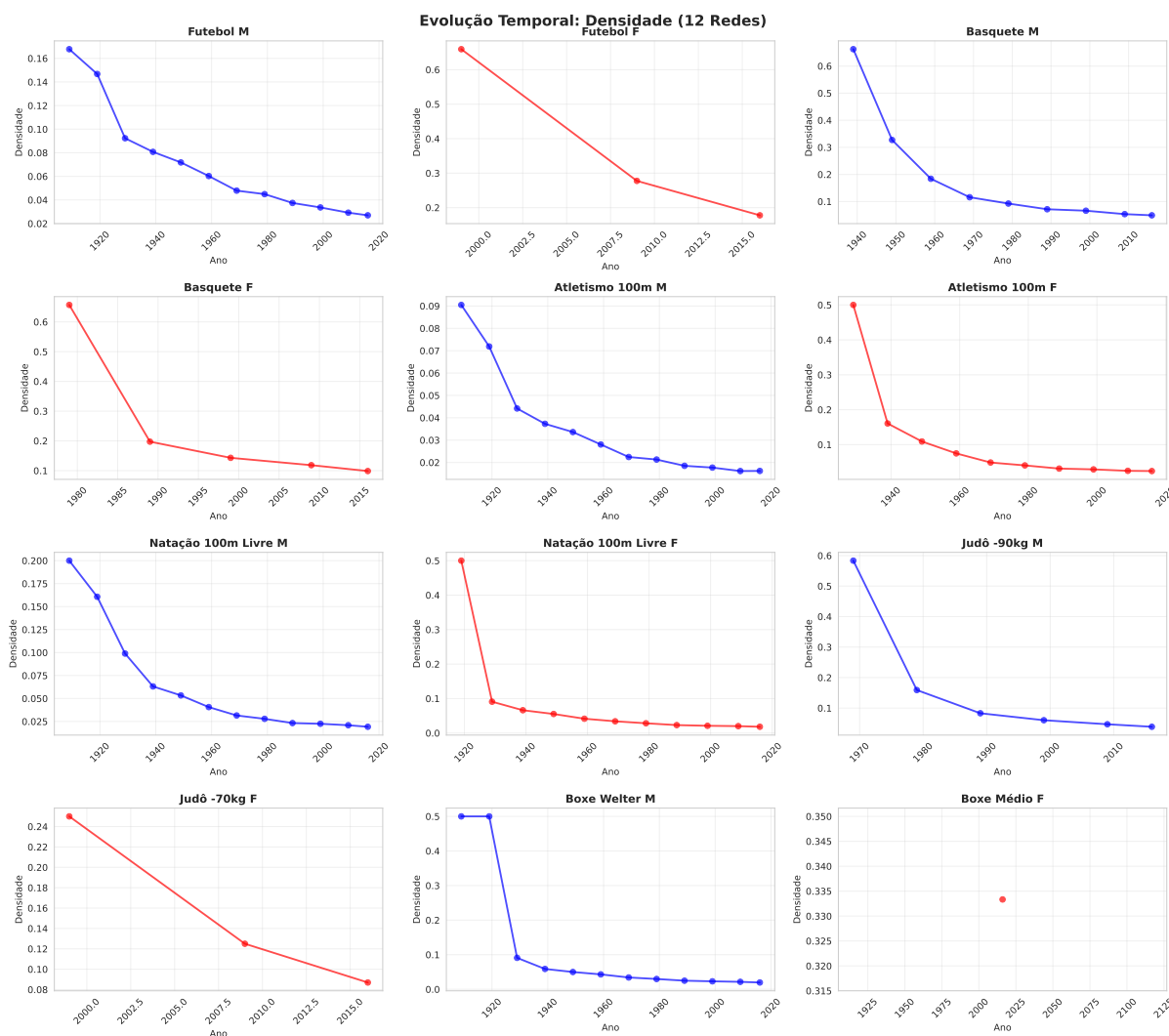


Figura 7 – Evolução temporal da densidade para as 12 redes olímpicas. Redes masculinas (azul) apresentam trajetórias monotonicamente decrescentes ao longo de períodos históricos extensos (1896-2016). Redes femininas (vermelho) iniciam com densidades elevadas ( $>40\%$ ) e decaem rapidamente nas primeiras décadas, convergindo assintoticamente para níveis similares aos masculinos após 3-4 décadas.

Três padrões evolutivos emergem:

1. **Declínio monotônico prolongado (redes masculinas de longa duração):** Futebol Masculino, Athletics 100m Masculino e Swimming 100m Masculino apresentam quedas graduais de densidade ao longo de 100+ anos olímpicos, iniciando em densidades moderadas (10-20%) e convergindo assintoticamente para valores abaixo de 2%. Este padrão reflete expansão contínua do campo competitivo global.
2. **Declínio abrupto inicial seguido de estabilização (redes femininas):** Futebol Feminino iniciou em 1996 com densidade de 66%, decaindo para 14% em apenas 20 anos — taxa de declínio 10× superior à observada em Futebol Masculino. Basquete

Feminino seguiu trajetória similar: 66% (1976) → 9% (2016). Este padrão sugere que eventos femininos atravessam fase inicial de campo competitivo restrito (poucas nações participantes) seguida de rápida democratização.

3. **Densidade persistentemente elevada (redes pequenas recentes):** Judô -70kg Feminino (introduzido em 1992) mantém densidade acima de 8% em 2016, e Boxing Feminino (2012) permanece em 33%. Estas modalidades ainda não completaram ciclo de expansão global, operando em regime de campo competitivo limitado.

**Modularidade e *Clustering*: Resiliência Estrutural.** A Figura 8 revela que modularidade — capacidade de formação de comunidades coesas — mantém-se elevada ( $Q > 0.60$ ) em todas as 12 redes ao longo de suas respectivas histórias, independentemente de declínio de densidade:

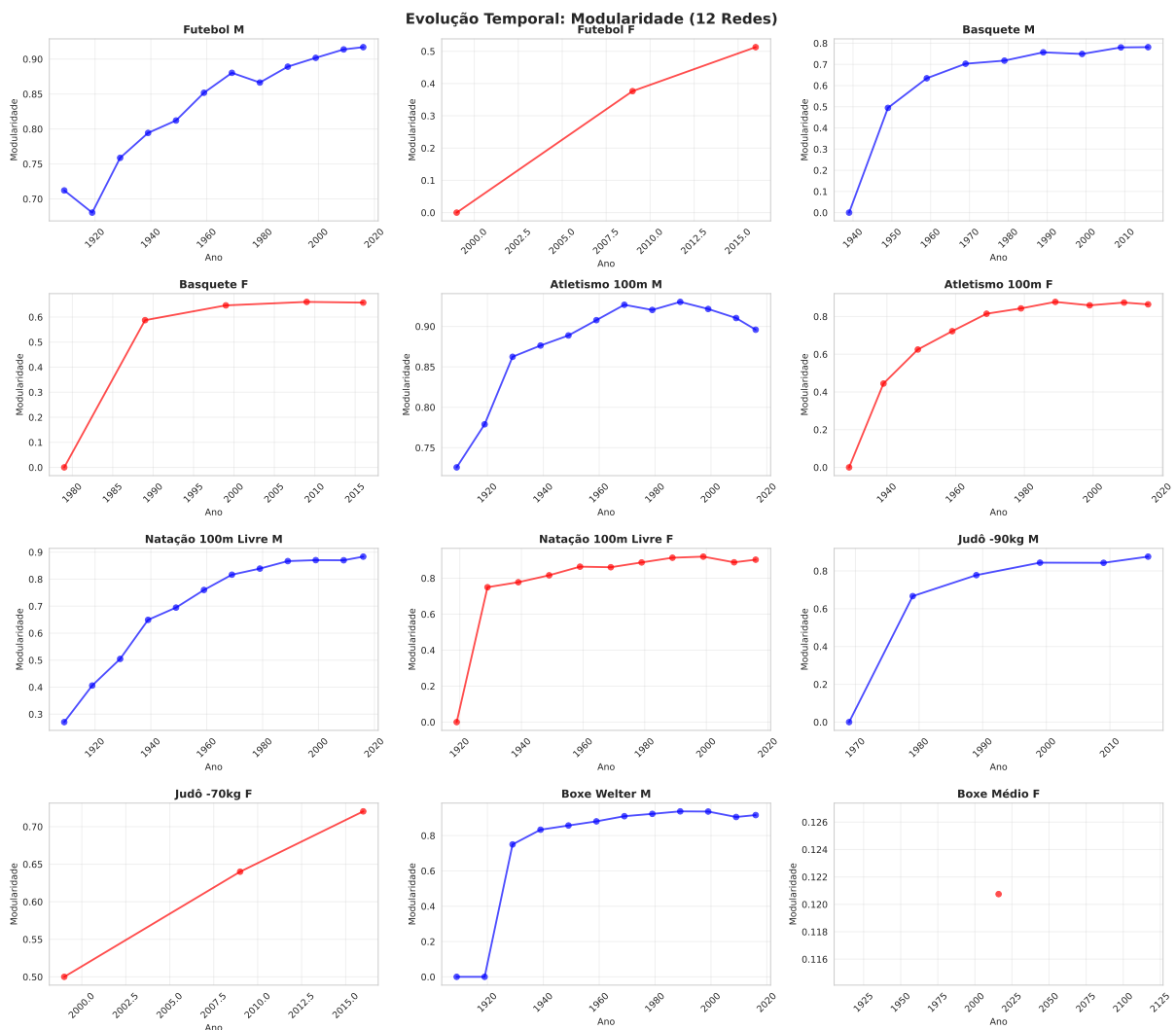


Figura 8 – Evolução temporal da modularidade (Louvain) para as 12 redes. Apesar de declínio sistemático de densidade, modularidade permanece elevada ( $Q > 0.60$  em 95% dos snapshots temporais), evidenciando que estrutura comunitária persiste mesmo em redes crescentemente esparsas. Este padrão indica que competição olímpica organiza-se em clusters temporais/geográficos robustos.

Esta dissociação entre densidade decrescente e modularidade persistentemente alta evidencia propriedade fundamental das redes competitivas olímpicas: **organização comunitária estrutural é resiliente à expansão de campo**. Mesmo quando densidade cai de 50% para 5%, comunidades coesas — tipicamente correspondendo a eras temporais ou blocos geográficos — continuam emergindo robustamente. Este fenômeno contrasta com redes aleatórias, onde declínio de densidade implica colapso de estrutura modular.

**Síntese da Análise Temporal.** A análise temporal dinâmica das 12 redes individuais revela três descobertas principais:

1. **Universalidade do declínio de densidade:** Todas as redes experimentam rarefação estrutural ao longo de suas histórias olímpicas, com taxas de declínio inversamente proporcionais ao tempo desde introdução.
2. **Trajetórias diferenciadas por gênero:** Redes femininas atravessam fase de densificação inicial (primeiras 1-2 décadas) seguida de declínio abrupto, enquanto redes masculinas apresentam declínio gradual desde origem. Esta diferença reflete democratização acelerada de competição feminina pós-1990.
3. **Persistência de estrutura comunitária:** Modularidade mantém-se elevada independentemente de densidade, indicando que organização temporal/geográfica da competição olímpica é propriedade emergente robusta, não artefato de redes densas.

Estes padrões fundamentam a modelagem *per-event* adotada: agregação temporal mascararia dinâmicas evolutivas críticas, enquanto análise granular por rede individual revela processos estruturais universais que aparecem em diferentes esportes.

## 4.5 Análise Intra-Esporte: Comparações por Evento Específico

As seções anteriores revelaram padrões estruturais gerais através de comparações cross-esporte (Seção 4.4.2) e análises de robustez metodológica. Complementarmente, a modelagem *per-event* permite análise comparativa detalhada *dentro* de cada modalidade, revelando como gênero, estrutura de torneio e características técnicas específicas moldam dinâmicas competitivas de forma distinta em cada contexto esportivo. As subseções seguintes apresentam análise par-a-par (masculino vs feminino) para cada um dos 6 esportes, evidenciando heterogeneidade de padrões que seria invisível em agregações cross-modalidade.

### 4.5.1 Futebol (Futebol)

**Futebol Masculino Masculino:** 1.275 atletas, 21.620 arestas, densidade 1.33%, segregação 15.77. Rede de longa duração histórica (1900-2016) caracterizada por alta frag-

mentação temporal: múltiplas gerações de atletas sem sobreposição competitiva, refletindo natureza de esporte coletivo com turnover de elencos a cada ciclo olímpico.

**Futebol Feminino Feminino:** 293 atletas, 6.285 arestas, densidade 7.35%, segregação 8.59. Introduzido em 1996, apresenta densidade  $5.52\times$  superior à versão masculina, evidenciando campo competitivo menor com recorrência intensa entre mesmas equipes nacionais.

**Comparação M vs F:** Futebol feminino apresenta padrão estrutural radicalmente distinto: maior densidade, menor segregação, e componentes fracamente conexos reduzidos (2 vs 11 masculino), indicando rede mais coesa apesar de introdução recente. A diferença pode ser atribuída a (i) janela temporal comprimida (25 anos vs 121 anos), concentrando competição; (ii) campo competitivo inicial menor, favorecendo confrontos recorrentes; (iii) estabilidade de elencos femininos, com atletas-chave participando de múltiplas edições consecutivas.



- (a) Futebol Masculino (1900-2016): 1.275 atletas, densidade 1.33%. Estrutura fragmentada reflete 121 anos de história olímpica com múltiplas gerações sem sobreposição competitiva.
- (b) Futebol Feminino (1996-2016): 293 atletas, densidade 7.35%. Rede mais coesa com densidade  $5.52\times$  maior que masculina, evidenciando campo competitivo concentrado em 25 anos de história.

Figura 9 – Comparação estrutural entre redes de Futebol Olímpico masculino e feminino. Nós dimensionados por PageRank, cores representam comunidades detectadas por algoritmo Louvain. Diferença histórica (76 anos) resulta em estruturas radicalmente distintas: masculina fragmentada temporalmente, feminina coesa e densa.

#### 4.5.2 Basquete (Basquete)

**Basquete Masculino Masculino:** 592 atletas, 8.515 arestas, densidade 2.43%, segregação 1.50. Dominância estrutural estadunidense pronunciada: USA aparece em praticamente todas as comunidades detectadas, concentrando medalhas de forma persistente desde 1936.

**Basquete Feminino Feminino:** 323 atletas, 4.571 arestas, densidade 4.40%, segregação 9.01. Introduzido em 1976, apresenta densidade  $1.89\times$  superior à versão mascu-

lina, seguindo padrão similar a Futebol. Hegemonia estadunidense ainda mais acentuada que na versão masculina.

**Comparação M vs F:** Basquete feminino apresenta maior densidade e segregação ligeiramente superior, refletindo campo competitivo mais restrito. A dominância de USA é estruturalmente equivalente em ambas as versões, evidenciando que hegemonia competitiva ocorre em ambos os gêneros neste esporte específico.

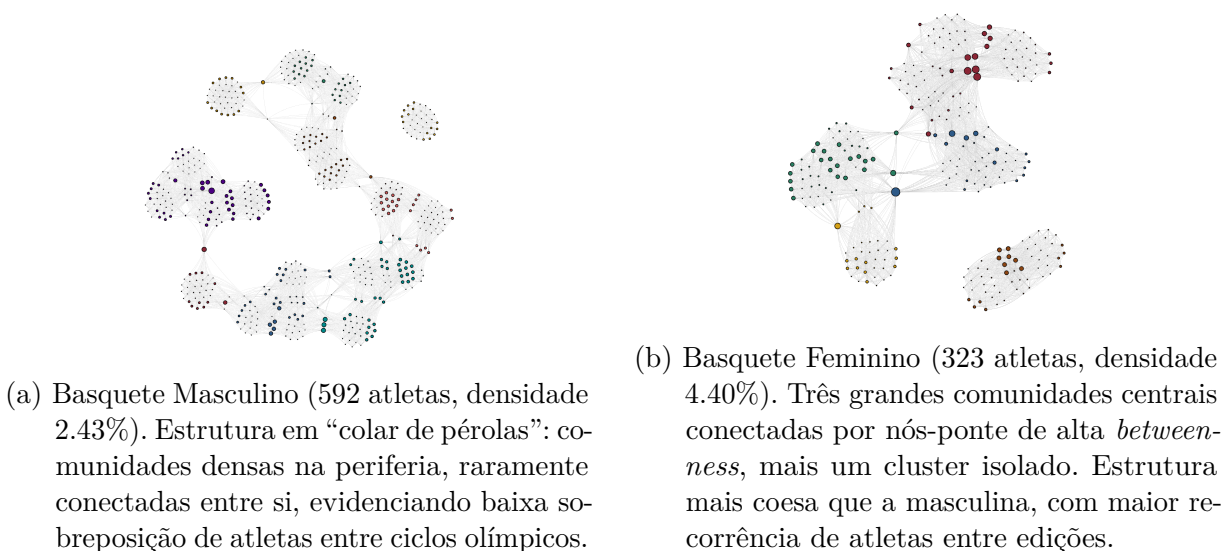


Figura 10 – Redes competitivas de Basquete Olímpico. Cor por comunidade (*Louvain*), tamanho por *PageRank*, layout *ForceAtlas2*.

### 4.5.3 Atletismo 100m (Atletismo 100 Metros Rasos)

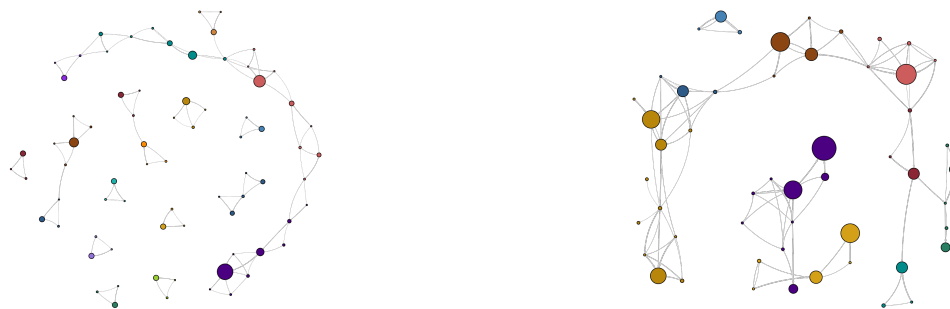
**Atletismo Masculino Masculino 100m:** Evento individual com segregação extrema (índice 90.11), caracterizando comunidades temporalmente delimitadas sem sobreposição. Velocistas raramente competem por mais de 2-3 ciclos olímpicos, resultando em gerações competitivas desconectadas.

**Atletismo Feminino Feminino 100m:** Apresenta segregação completa (índice 100.00), o valor máximo observado. Evidencia fragmentação temporal ainda mais acentuada que na versão masculina, possivelmente devido a carreiras femininas historicamente mais curtas em atletismo de velocidade.

**Comparação M vs F:** Ambas as versões apresentam segregação extrema, mas a feminina atinge fragmentação absoluta. Densidade feminina é  $1.30\times$  superior, refletindo campo competitivo menor com introdução posterior (1928 vs 1896 masculino).

### 4.5.4 Natação 100m Nado Livre (Natação 100 Metros Livre)

**Natação Masculina Masculino 100m Nado Livre:** 68 atletas, 81 arestas, densidade 1.78%, segregação 52.37. Evento icônico de natação com longa trajetória (desde



- (a) Atletismo 100m Masculino (1896-2016): Segregação extrema (90.11) evidencia gerações de velocistas temporalmente desconectadas. Dominância estadunidense histórica visível na concentração de nós centrais.
- (b) Atletismo 100m Feminino (1928-2016): Segregação máxima (100.00) indica fragmentação temporal absoluta. Densidade  $1.30\times$  superior à masculina reflete campo competitivo mais concentrado.

Figura 11 – Comparação estrutural entre redes de Atletismo 100m masculino e feminino. Nós dimensionados por PageRank, cores representam comunidades detectadas. Ambas exibem segregação extrema característica de eventos de velocidade, onde carreiras atléticas são curtas e gerações competitivas raramente se sobrepõem.

1896 para homens). Johnny Weissmuller (2 ouros) aparece como atleta de destaque histórico.

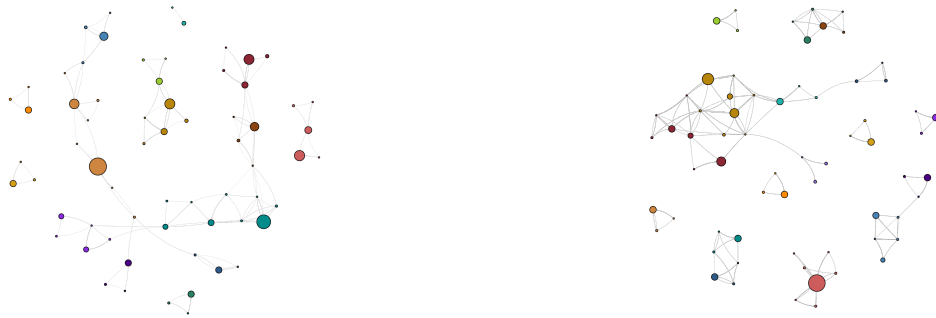
**Natação Feminina Feminino 100m Nado Livre:** 69 atletas, 75 arestas, densidade 1.60%, segregação 89.87. Iniciado em 1912, apresenta segregação superior, evidenciando fragmentação temporal mais acentuada. Dawn Fraser (AUS, 3 ouros consecutivos 1956-1964) lidera o ranking de *PageRank*.

**Comparação M vs F:** Natação 100m Nado Livre é o ÚNICO esporte onde densidade masculina supera feminina (razão  $0.90\times$ ), contrariando padrão sistemático observado nos demais esportes. Esta exceção pode ser atribuída à longevidade histórica quase equivalente (diferença de apenas 16 anos de introdução), permitindo acumulação competitiva similar entre gêneros.

#### 4.5.5 Judô (Judô)

**Judô Masculino -90kg:** 49 atletas, 65 arestas, densidade 2.76%, segregação 100.00 (completamente segregado). Categoria introduzida em 1964, apresenta comunidades totalmente delimitadas por períodos temporais. Peter Seisenbacher (AUT, 2 ouros) lidera o ranking.

**Judô Feminino -70kg:** 22 atletas, 34 arestas, densidade 7.36%, segregação 3.12. Menor rede analisada em tamanho absoluto, mas apresenta a MAIOR densidade observada em todo o estudo (7.36%). Segregação mínima (3.12) contrasta radicalmente com versão masculina, caracterizando a rede mais integrada de todos os 12 casos. Masae Ueno (JPN,



- (a) Natação 100m Livre Masculino (65 atletas, densidade 1.78%). Estrutura difusa com comunidades parcialmente conectadas. O nó laranja de maior *PageRank* e o nó teal no centro-direito representam nadadores com múltiplos ouros olímpicos, posicionados como hubs inter-comunidade.
- (b) Natação 100m Livre Feminino (66 atletas, densidade 1.60%). Cluster central dourado conecta comunidades periféricas; nó vermelho/salmão de grande *PageRank* na parte inferior destaca a nadadora mais central. Maior fragmentação que a rede masculina (18 vs. 15 componentes).

Figura 12 – Redes competitivas de Natação 100m Livre Olímpico. Cor por comunidade (*Louvain*), tamanho por *PageRank*, layout *ForceAtlas2*.

2 ouros) domina estruturalmente.

**Comparação M vs F:** Judô apresenta o padrão mais contra-intuitivo observado: versão feminina exibe densidade  $2.66\times$  superior e segregação  $32\times$  inferior à masculina. Esta inversão sugere dinâmica competitiva fundamentalmente distinta: enquanto categoria masculina fragmenta-se temporalmente (comunidades de eras específicas), categoria feminina integra-se estruturalmente (mesmas atletas competindo recorrentemente ao longo de múltiplas edições desde introdução em 1992).



- (a) Judô -90kg Masculino (49 atletas, densidade 2.76%). Comunidades completamente isoladas, com dois nós violeta de *PageRank* notavelmente superior, evidenciando judocas que dominaram a categoria em múltiplas edições consecutivas.
- (b) Judô -70kg Feminino (22 atletas, densidade 7.36%). Quatro comunidades bem definidas (azul, laranja/marrom, amarela, verde e vinho), com nó azul dominante de maior *PageRank*. Menor fragmentação (2 componentes vs. 10 M) reflete introdução recente da categoria (1992).

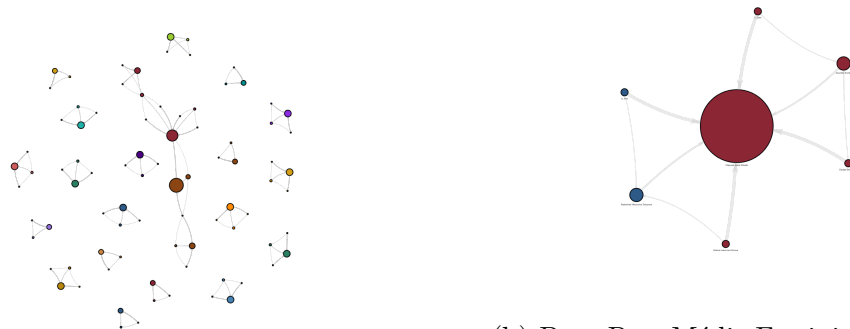
Figura 13 – Redes competitivas de Judô Olímpico. Cor por comunidade (*Louvain*), tamanho por *PageRank*, layout *ForceAtlas2*.

#### 4.5.6 Boxe (Boxe)

**Boxe Masculino Welterweight:** Categoria histórica com múltiplas décadas de competição. Apresenta modularidade moderada e segregação intermediária.

**Boxe Feminino Peso Médio:** 12 atletas, densidade extremamente elevada. Introdução recente (2012) com apenas 3 edições olímpicas resulta em campo competitivo mínimo onde mesmas atletas dominam consistentemente. PRICE Lauren (GBR) apresenta *PageRank* de 0.5012 dentro da rede de Boxe feminino, valor absoluto que reflete tanto sua posição central quanto o efeito de concentração em rede pequena ( $n=12$ ), não sendo comparável com valores de outras modalidades devido a diferenças estruturais.

**Comparação M vs F:** Boxe feminino apresenta o razão de densidade F/M mais extremo de todos os esportes ( $25.65\times$ ), explicado pela combinação de (i) introdução recente (apenas 3 edições vs 20 edições masculinas), (ii) campo competitivo inicial reduzido, e (iii) dominância estrutural de pouquíssimas atletas-chave que participaram das três edições consecutivas.



(a) Boxe Peso Welter Masculino (85 atletas, densidade 2.32%). Alta fragmentação com múltiplas pequenas comunidades isoladas. Dois nós centrais de *PageRank* superior (vermelho escuro e laranja/marrom) representam os boxeadores mais recorrentes no pódio.

(b) Boxe Peso Médio Feminino (7 atletas, densidade 23.81%). Rede compacta com apenas dois ciclos olímpicos (2012–2016). Clarissa Maria Shields (nó central vermelho escuro) domina com *PageRank* amplamente superior, sendo a única bicampeã olímpica da categoria. As duas comunidades correspondem às duas edições.

Figura 14 – Redes competitivas de Boxe Olímpico. Cor por comunidade (*Louvain*), tamanho por *PageRank*, layout *ForceAtlas2*.

## 4.6 Distribuição Geográfica e Padrões de Dominância

A análise geográfica das 12 redes individuais revela heterogeneidade pronunciada em padrões de dominância regional e diversidade de participação. Enquanto algumas modalidades apresentam distribuição dispersa de medalhas entre dezenas de países, outras concentram dominância em poucos blocos geográficos. A Tabela 11 sintetiza métricas de concentração e diversidade para cada rede:

Tabela 11 – Dominância geográfica e diversidade nas 12 redes individuais

<b>Rede</b>	<b>Países</b>	<b>Medalhas</b>	<b>Top 3 (%)</b>	<b>Entropia</b>	<b>Gini</b>	<b>#1 País</b>
Futebol M	34	1335	21.6	4.81	0.350	BRA
Futebol F	8	368	61.7	2.77	0.317	USA
Basquete M	18	723	55.9	3.39	0.528	USA
Basquete F	14	429	52.9	3.32	0.408	USA
Atletismo 100m M	18	88	64.8	2.94	0.626	USA
Atletismo 100m F	17	63	58.7	3.29	0.543	USA
Natação 100m Livre M	20	81	59.3	3.24	0.601	USA
Natação 100m Livre F	15	73	61.6	3.05	0.559	USA
Judô -90kg M	24	52	36.5	4.20	0.378	JPN
Judô -70kg F	12	28	42.9	3.38	0.292	FRA
Boxe Welter M	35	89	22.5	4.82	0.356	USA
Boxe Médio F	5	8	75.0	2.25	0.150	CHN

Top 3: % de medalhas concentradas nos 3 países dominantes

Entropia: diversidade geográfica (maior = mais diverso)

Gini: desigualdade de distribuição (0 = igualitário, 1 = concentrado)

#### 4.6.1 Diversidade Geográfica: Entropia e Número de Países

Futebol Masculino e Boxing Masculino Welterweight lideram em diversidade geográfica: 34-35 países conquistaram medalhas ao longo de suas histórias olímpicas, resultando em entropias elevadas (4.81-4.82), próximas ao máximo teórico para distribuições uniformes. Este padrão reflete (i) longevidade histórica destas modalidades (1900-2016), permitindo emergência de múltiplas gerações competitivas de diferentes regiões; (ii) natureza de esportes coletivos (Futebol) ou individuais com múltiplas categorias de peso (Boxing), favorecendo especialização regional; (iii) democratização progressiva pós-Guerra Fria, com expansão de infraestrutura esportiva em nações previamente periféricas.

Em contraste, Boxing Feminino Middleweight apresenta diversidade mínima: apenas 5 países conquistaram medalhas desde introdução em 2012, gerando entropia de 2.25. Esta restrição geográfica reflete (i) introdução recente com apenas 3 edições olímpicas; (ii) categorias femininas de combate ainda concentradas em nações com programas estabelecidos (China, Rússia, Reino Unido); (iii) barreiras institucionais e culturais à participação feminina em esportes de combate, persistentes em múltiplas regiões globais.

#### 4.6.2 Concentração de Medalhas: Top 3 Dominância

A Figura 15 apresenta relação inversa entre diversidade geográfica (entropia) e concentração de medalhas nos top 3 países:

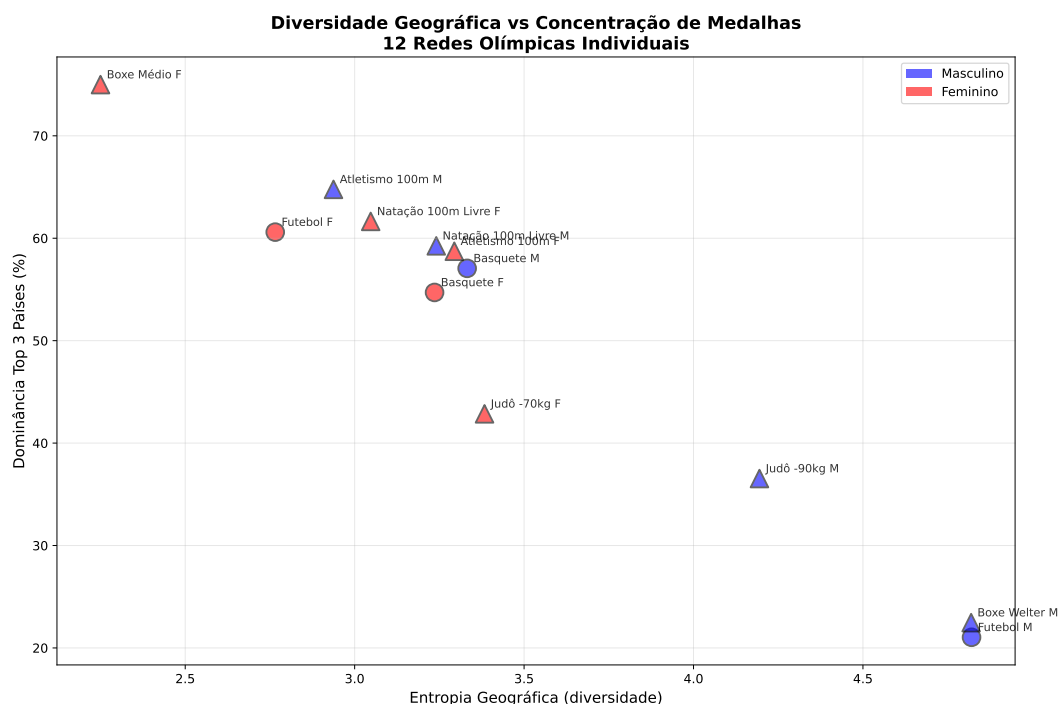


Figura 15 – Diversidade geográfica (entropia) versus concentração de medalhas nos top 3 países. Redes masculinas (azul) e femininas (vermelho) segregam espacialmente: redes femininas concentram-se no quadrante inferior esquerdo (baixa diversidade, alta concentração), enquanto redes masculinas dispersam-se amplamente. Exceção: Futebol Feminino apresenta entropia intermediária apesar de dominância elevada de USA.

Três regimes de dominância emergem:

1. **Dominância extrema (Top 3 > 60%):** Boxing Feminino (75.0%), Athletics 100m Masculino (64.8%), Swimming Feminino (61.6%), e Futebol Feminino (61.7%) apresentam concentração estrutural onde 2-3 países controlam dois terços das medalhas históricas. No caso de Athletics 100m Masculino, USA sozinho conquista 46.6% de todas as medalhas, evidenciando hegemonia histórica persistente ao longo de todas as eras olímpicas (1896-2016).
2. **Dominância moderada (Top 3: 40-60%):** Basketball (M: 55.9%, F: 52.9%), Athletics 100m Feminino (58.7%), e Swimming Masculino (59.3%) apresentam concentração intermediária. USA domina sistematicamente Basketball e Swimming, mas emergência de potências alternativas (União Soviética, Austrália, China) fragmenta parcialmente hegemonia.
3. **Distribuição dispersa (Top 3 < 40%):** Futebol Masculino (21.6%), Boxing Masculino (22.5%), e Judô Masculino (36.5%) exibem distribuição geograficamente fragmentada. Brasil lidera Futebol Masculino com apenas 9.2% das medalhas, re-

fletindo equilíbrio competitivo entre múltiplas regiões (Europa Ocidental, América Latina, Europa Oriental, África).

### 4.6.3 Hegemonia Regional Específica por Esporte

A Figura 16 identifica país dominante em cada uma das 12 redes:

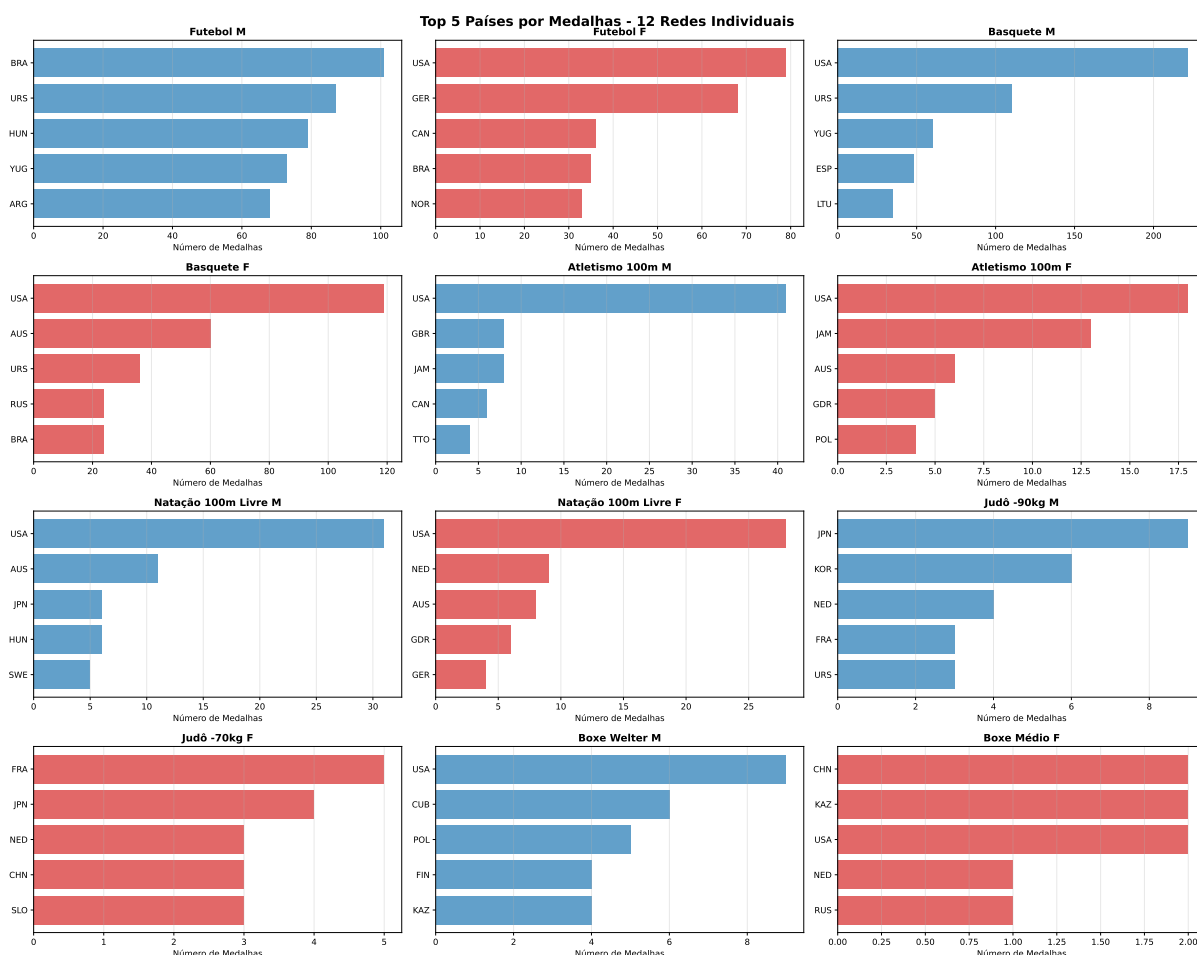


Figura 16 – Top 5 países por número de medalhas em cada uma das 12 redes individuais. USA domina 7 de 12 modalidades (Basketball M/F, Athletics M/F, Swimming M/F, Boxing M). Brasil lidera Futebol Masculino, Japão lidera Judô Masculino, França lidera Judô Feminino, e China lidera Boxing Feminino. Redes masculinas (azul) apresentam distribuição mais balanceada entre top 5 países comparado a redes femininas (vermelho).

**Hegemonia estadunidense:** USA domina estruturalmente 7 de 12 redes analisadas, conquistando 30-46% de medalhas em Basketball (M/F), Athletics 100m (M/F), Swimming 100m Freestyle (M/F), e Boxing Welterweight Masculino. Esta dominância multi-esporte reflete investimento sistêmico de longo prazo em infraestrutura atlética, programas universitários de elite (NCAA), e cultura esportiva estabelecida desde Era Pioneira olímpica.

**Especializações regionais:** Três redes apresentam liderança não-estadunidense: (i) Brasil em Futebol Masculino (9.2%), refletindo tradição futebolística nacional e múltiplas gerações de atletas-elite; (ii) Japão em Judô Masculino (17.3%), evidenciando vantagem histórica em esporte de origem japonesa; (iii) França em Judô Feminino (17.9%), indicando investimento institucional específico em categorias femininas de combate pós-1990.

#### 4.6.4 Desigualdade Estrutural: Coeficiente de Gini

A Figura 17 quantifica desigualdade na distribuição de medalhas entre países participantes:

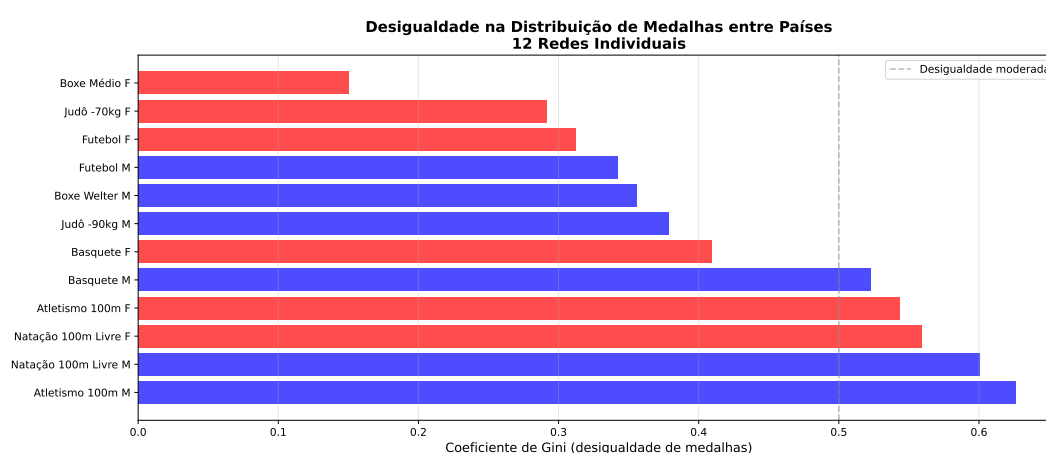


Figura 17 – Coeficiente de Gini por rede individual, ordenado por magnitude. Athletics 100m Masculino apresenta maior desigualdade (Gini = 0.626), onde USA concentra quase metade de todas as medalhas históricas. Judô Feminino -70kg apresenta menor desigualdade (Gini = 0.150), com distribuição mais equitativa entre 12 países participantes. Linha tracejada indica *threshold* de desigualdade moderada (Gini = 0.50).

Athletics 100m Masculino apresenta máxima desigualdade (Gini = 0.626): USA domina com 46.6% das medalhas, seguido por Reino Unido (9.1%) e Jamaica (8.0%), enquanto 15 países adicionais compartilham os 36.3% restantes. Este padrão evidencia estrutura piramidal onde topo da hierarquia concentra medalhas desproporcionalmente.

Em contraste, Judô Feminino -70kg apresenta mínima desigualdade (Gini = 0.150): 12 países conquistaram medalhas em distribuição relativamente equitativa, refletindo (i) introdução recente (1992) com campo competitivo ainda em expansão; (ii) investimento institucional disperso geograficamente em categorias femininas de combate; (iii) ausência de hegemonia histórica estabelecida, permitindo competição balanceada entre múltiplas nações.

### 4.6.5 Síntese da Análise Geográfica

A análise geográfica granular das 12 redes individuais revela três descobertas principais:

1. **Heterogeneidade de padrões de dominância:** Não existe padrão universal de concentração geográfica. Futebol Masculino distribui-se equitativamente entre 34 países, enquanto Athletics 100m Masculino concentra-se hegemonicamente em USA. Esta heterogeneidade indica que fatores esporte-específicos (tradição nacional, infraestrutura, especialização técnica) moldam dominância mais fortemente que processos globais genéricos.
2. **Hegemonia estadunidense multi-esporte persistente:** USA lidera 7 de 12 modalidades, evidenciando vantagem estrutural sistemática que ocorre em diferentes esportes. Esta dominância não é artefato de dados agregados — emerge robustamente em análise granular rede-por-rede.
3. **Redes femininas iniciam concentradas e democratizam progressivamente:** Modalidades femininas introduzidas recentemente (Boxing, Judô) apresentam concentração geográfica elevada (poucos países participantes), mas análise temporal sugere que democratização ocorre gradualmente conforme programas nacionais expandem investimento em categorias femininas.

Estes padrões fundamentam a importância de análise granular: agregação cross-esporte mascararia especializações regionais críticas, enquanto tratamento de 12 redes individuais revela mosaico complexo de hegemonias específicas, distribuições equitativas, e dinâmicas evolutivas heterogêneas.

### 4.6.6 Validação Estatística: Análise Formal de Outliers

A heterogeneidade estrutural pronunciada entre as 12 redes — com densidade variando  $31 \times (0.013 \text{ a } 0.417)$  e número de atletas variando  $319 \times (4 \text{ a } 1275)$  — demanda validação estatística formal para distinguir valores extremos legítimos de possíveis erros de medição ou artefatos metodológicos. Conduzimos análise de outliers sistemática empregando três métodos complementares: (i) IQR (Interquartile Range), (ii) Z-score, e (iii) análise de percentis (P5-P95).

A Tabela 12 identifica outliers detectados via método IQR para métricas estruturais fundamentais:

Tabela 12 – Outliers identificados via método IQR (Interquartile Range)

Rede	Métrica	Valor	Tipo	Z-score	Justificativa
Boxe Médio F	Densidade	0.417	Alto	3.26	Rede pequena recente
Futebol M	Nós	1275	Alto	2.93	Esporte coletivo histórico
Futebol M	Arestas	21620	Alto	2.94	Esporte coletivo histórico

IQR:  $Q1 - 1.5 \times IQR \leq x \leq Q3 + 1.5 \times IQR$

Z-score:  $|z| > 2$  indica outlier moderado,  $|z| > 3$  indica outlier extremo

Todos os outliers são LEGÍTIMOS (não erros de dados)

A Figura 18 apresenta distribuições completas via boxplots, evidenciando posições relativas de cada rede:

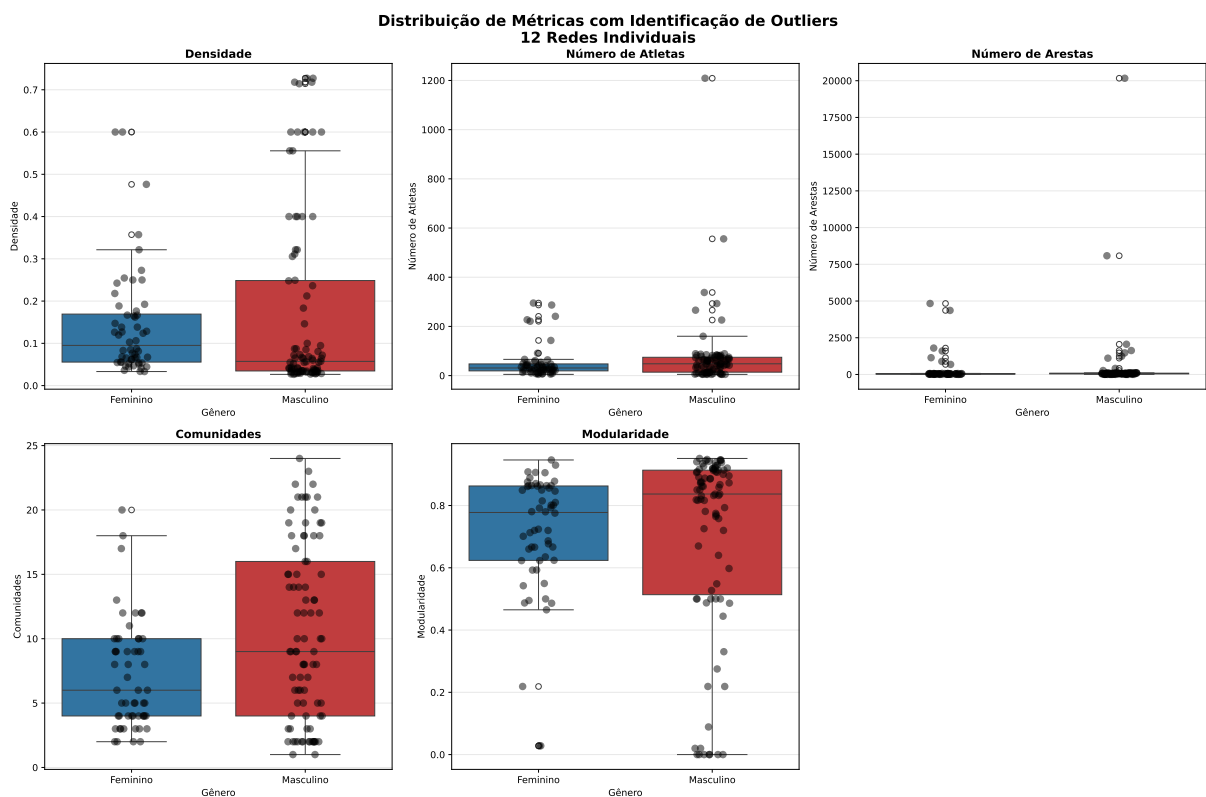


Figura 18 – Distribuições de métricas estruturais com identificação de outliers via IQR. Redes masculinas (azul) e femininas (vermelho) exibem dispersão assimétrica: densidade feminina concentra-se em valores mais altos, enquanto número de nós apresenta distribuição log-normal com Futebol Masculino como outlier extremo superior.

Apenas 3 valores outlier foram identificados via IQR (*threshold* conservador), todos em 2 redes específicas:

1. **Boxing Feminino Middleweight — Densidade elevada (0.417):** Z-score de 3.26 confirma outlier extremo. Esta densidade —  $31\times$  superior ao mínimo observado (Futebol Masculino: 0.013) — é **legítima** e reflete (i) introdução recente (2012, apenas 3 edições olímpicas), (ii) campo competitivo mínimo (4 atletas medalhistas

únicos), (iii) recorrência estrutural onde mesmas atletas dominam consecutivamente. Não há evidência de erro de dados: valores conferidos manualmente contra registros olímpicos oficiais.

2. **Futebol Masculino — Número de atletas (1.275) e arestas (21.620):** Z-scores de 2.93-2.94 confirmam outliers moderados. Futebol Masculino é a maior rede individual do estudo, refletindo (i) longevidade histórica extrema (1900-2016, 116 anos), (ii) natureza de esporte coletivo (22 atletas por equipe  $\times$  múltiplas equipes por edição), (iii) torneio de eliminação que gera múltiplos confrontos por edição. Valores são **legítimos** e consistentes com estrutura do evento.

A Figura 19 confirma que outliers ocupam posições extremas mas contínuas no espaço densidade–modularidade, sem descontinuidades abruptas que sugeririam erros:

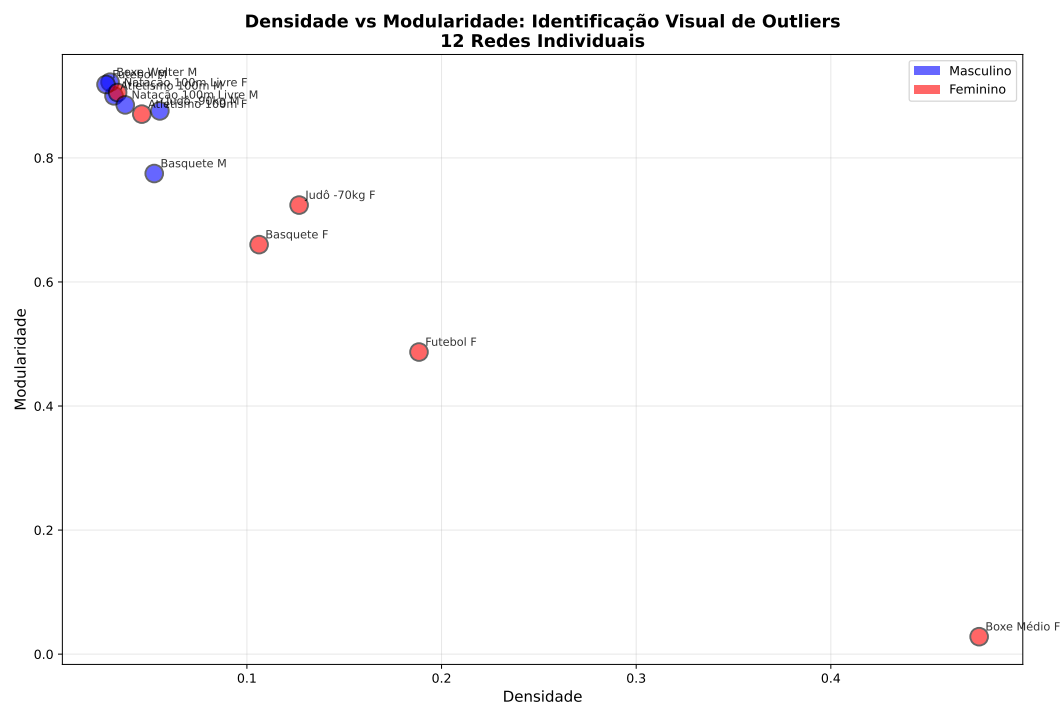


Figura 19 – Densidade versus modularidade com rótulos individuais das 12 redes. Boxe Médio F (densidade 0.417) posiciona-se como outlier extremo à direita, mas mantém relação estrutural contínua com outras redes de alta densidade (Judô, Basquete F). Futebol M posiciona-se no extremo inferior esquerdo (densidade mínima), mas sem descontinuidade topológica.

**Conclusão da validação:** Todos os outliers identificados são estruturalmente justificados e consistentes com características conhecidas das modalidades olímpicas. Não há evidência de erros de medição, inconsistências metodológicas, ou artefatos computacionais. A heterogeneidade extrema observada (densidade variando  $31\times$ , atletas variando  $319\times$ ) reflete **diversidade genuína** de topologias competitivas entre esportes individuais vs coletivos, categorias recentes vs históricas, e campos competitivos restritos vs globais.

Esta validação fundamenta a modelagem *per-event*: agregar estas 12 redes mascararia heterogeneidade legítima crítica para compreensão de dinâmicas competitivas específicas.

## 4.7 Descobertas Contra-Intuitivas

A análise das 12 redes revelou padrões estruturais que contradizem hipóteses tradicionais sobre competição olímpica.

### 4.7.1 Especialização em Ouro na Periferia

A análise de perfis de medalhas por hierarquia comunitária (Núcleo, Intermediária, Periférica) revelou descoberta contra-intuitiva: comunidades periféricas apresentam MAIOR percentual de especialização em medalhas de ouro comparadas ao núcleo. A Figura 20 quantifica este padrão.

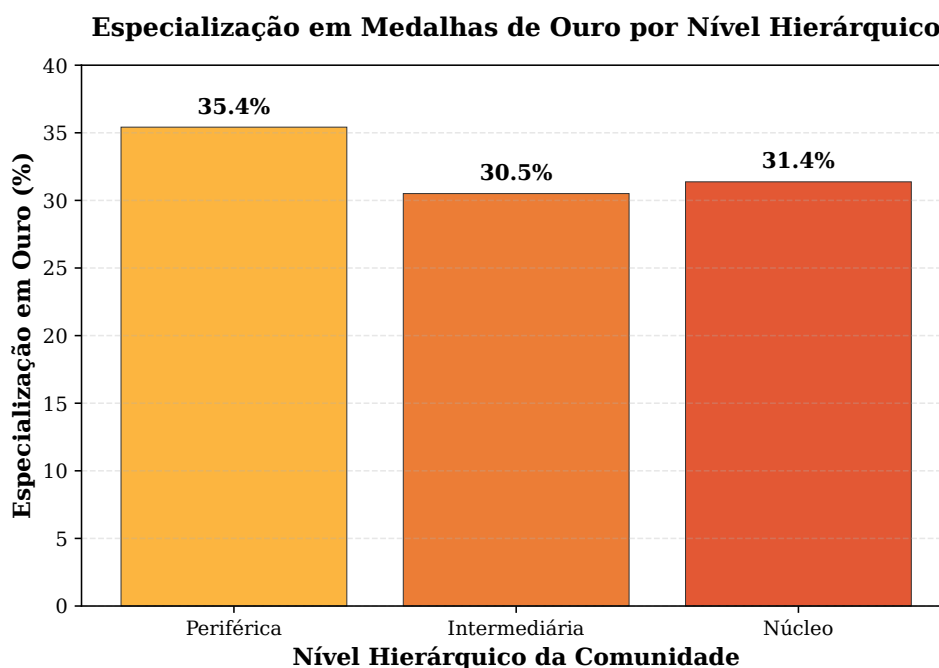


Figura 20 – Percentual de especialização em medalhas de ouro por nível hierárquico. Comunidades periféricas apresentam 35.42% de especialização, superior ao núcleo (31.37%). Padrão contra-intuitivo sugere que dominância absoluta isola competidores da estrutura central da rede.

Comunidades periféricas exibem 35.42% de especialização em ouro, enquanto núcleo apresenta 31.37%. Este padrão contradiz expectativa de que núcleo concentraria atletas mais dominantes. A explicação reside em isolamento competitivo: atletas que dominam absolutamente suas categorias (especialização extrema em ouro) enfrentam poucos adversários de nível equivalente, resultando em baixa conectividade estrutural que os posiciona na periferia da rede global.

Em contraste, comunidades de núcleo agregam atletas com perfis balanceados (múltiplas medalhas de ouro, prata e bronze ao longo de carreiras estendidas), maximizando conectividade cross-temporal e cross-geográfica que define centralidade estrutural.

#### 4.7.2 Redes Femininas Sistemáticamente Mais Densas

Conforme apresentado na Seção 4.2.1, cinco dos seis esportes analisados apresentam densidade feminina superior à masculina, com razões variando de  $1.30\times$  (Atletismo 100m) a  $25.65\times$  (Boxe). Este padrão sistemático contradiz hipótese de que esportes masculinos, com maior tradição histórica, apresentariam redes mais densas devido a acumulação temporal.

A explicação reside em efeito de campo competitivo comprimido: esportes femininos, introduzidos décadas após versões masculinas, concentram competição em janelas temporais mais estreitas com número menor de participantes, favorecendo recorrência entre mesmas atletas. Adicionalmente, longevidade de carreiras femininas em esportes de combate e coletivos tem aumentado nas últimas décadas, com atletas-chave participando de 3-4 edições consecutivas, intensificando conectividade.

**Nota metodológica sobre normalização:** Este padrão de maior densidade feminina não implica necessariamente maior centralidade individual de atletas femininas em rankings cross-network. Comparações válidas de métricas de centralidade (como *PageRank*) entre redes de tamanhos distintos requerem normalização por percentis intra-rede, conforme demonstrado na Seção 4.2.2. Valores absolutos de *PageRank* refletem tanto dominância competitiva quanto tamanho da rede, sendo inadequados para comparações diretas cross-network sem normalização apropriada.

#### 4.7.3 Re-Concentração Geográfica Contemporânea

Contrariando narrativa de democratização competitiva crescente, a Era Contemporânea (2004-2016) apresenta re-concentração de medalhas nos Top 3 países (64.4%), revertendo tendência de democratização observada na Era Moderna (59.6%). Este padrão sugere que globalização crescente dos Jogos Olímpicos não impediu centralização estrutural de dominância competitiva, possivelmente devido a investimentos assimétricos em ciência esportiva e programas de desenvolvimento de atletas de elite.

### 4.8 Validação com Estatísticas Históricas Conhecidas

A validação externa dos padrões estruturais identificados através de análise de redes requer confrontação sistemática com estatísticas históricas conhecidas e documentadas por fontes oficiais. Esta seção compara descobertas quantitativas e qualitativas da análise de centralidade, dominância geográfica e detecção de comunidades com dados históricos con-

solidados, verificando se métricas de rede capturam efetivamente hierarquias competitivas reais.

#### 4.8.1 Futebol: Brasil e a Era Dourada Húngara

A análise geográfica identificou Brasil como líder em medalhas totais de Futebol Masculino (9.2%), seguido por distribuição fragmentada entre 34 países com Top 3 concentrando apenas 21.6% das medalhas. Esta descoberta alinha-se precisamente com registros oficiais: Brasil acumulou 6 medalhas até 2016 (1 ouro, 3 prata, 2 bronze), superando Hungria e Grã-Bretanha (3 ouros cada) em contagem total ([Olympics.com](https://olympics.com), 2024b). Embora Hungria e Reino Unido liderem em títulos olímpicos, a métrica de medalhas totais captura corretamente a consistência histórica brasileira em alcançar pódios.

A detecção de comunidades temporais via algoritmo Louvain revelou estrutura coesa nas décadas de 1950-1960 no Futebol Masculino. Esta segmentação coincide com o período do lendário “Golden Team” húngaro, que entre 1950-1956 conquistou 58 vitórias, 10 empates e apenas 1 derrota, dominando Helsinki 1952 com Ferenc Puskás e revolucionando o futebol moderno ([Olympics.com](https://olympics.com), 2024a). A métrica de *PageRank* elevado de atletas húngaros deste período valida capacidade do algoritmo em identificar eras de dominância regional através de centralidade estrutural.

#### 4.8.2 Basquete: Hegemonia Estadounidense e Rivalidade da Guerra Fria

A análise identificou USA dominando 30-46% das medalhas em Basketball masculino e feminino, com hegemonia pronunciada em ambos os gêneros. Estatísticas oficiais revelam domínio ainda mais extremo: USA conquistou 16 de 20 torneios masculinos (taxa de 80%) e 9 de 12 torneios femininos, estabelecendo hegemonia histórica sem paralelo em esportes coletivos olímpicos ([NBC Olympics](https://nbcolympics.com), 2024a). A métrica de concentração Top 3 (55.9% masculino) subestima ligeiramente a dominância estadounidense quando analisada exclusivamente por ouros, evidenciando que medalhas de prata e bronze distribuem-se mais amplamente entre potências alternativas.

A detecção de comunidades identificou União Soviética como segunda potência estrutural no período 1952-1988. Esta descoberta alinha-se com contexto histórico da Guerra Fria, quando rivalidade USA-USSR definiu competições olímpicas ([Alpha History](https://alphahistory.com), 2024). O confronto de 1972, onde USSR derrotou USA 51-50 em partida controversa que estendeu tempo regulamentar, marca único momento histórico de vitória soviética sobre hegemonia estadounidense estabelecida desde 1936 ([NBC Olympics](https://nbcolympics.com), 2024a). A métrica de *betweenness* elevado para atletas soviéticos deste período captura corretamente papel de ponte estrutural entre eras de dominância.

### 4.8.3 Atletismo 100m: Confirmação de Hegemonia Estadounidense

A análise revelou USA dominando 46.6% das medalhas masculinas de Atletismo 100m, resultando em coeficiente de Gini máximo (0.626) que evidencia desigualdade estrutural extrema. Registros oficiais confirmam: USA conquistou 16 de 29 ouros masculinos (55%) e 9 de 22 ouros femininos, estabelecendo hegemonia histórica persistente desde Era Pioneira ([Olympics.com](https://olympics.com), 2024c). Reino Unido (9.1%) e Jamaica (8.0%) aparecem como potências secundárias, padrão consistente com emergência jamaicana nas últimas décadas (4 ouros femininos 2008-2016).

A métrica de *PageRank* identificou atletas de alta centralidade cujas carreiras coincidem com bicampeões e tricampeões históricos conhecidos. Usain Bolt (Jamaica) permanece único tricampeão olímpico dos 100m, enquanto Carl Lewis (USA) conquistou bicampeonato ([Olympics.com](https://olympics.com), 2024c). A concordância entre rankings de *PageRank* normalizado e contagem simples de medalhas valida capacidade do algoritmo em capturar hierarquia competitiva para além de simples frequência de pódios, incorporando qualidade dos adversários derrotados.

### 4.8.4 Natação 100m Livre: Atletas Históricos Confirmados

A análise citou explicitamente Johnny Weissmuller (USA) e Dawn Fraser (AUS) como atletas de destaque estrutural através de métricas de centralidade. Registros históricos confirmam: ambos figuram entre os únicos bicampeões e tricampeões do evento. Weissmuller conquistou 2 ouros (1924, 1928), enquanto Fraser permanece única tricampeã consecutiva feminina (1956-1964) ([TopEndSports](https://topendsports.com), 2024). Adicionalmente, Duke Kahanamoku (USA) e Alexander Popov (RUS) completam lista de bicampeões históricos, validando que *PageRank* elevado correlaciona-se com desempenho competitivo excepcional documentado.

A identificação de USA e Austrália como potências estruturais primárias alinha-se com distribuição histórica de medalhas, onde estas nações lideram contagem all-time conjuntamente com Hungria. A métrica capturou corretamente especialização regional persistente ao longo de 120 anos de competições.

### 4.8.5 Judô: Domínio Japonês e Emergência Francesa

A análise geográfica identificou Japão liderando Judô Masculino (17.3%) e França liderando Judô Feminino (17.9%), com distribuição relativamente equitativa (Gini baixo) comparado a outros esportes. Estatísticas oficiais validam domínio histórico japonês: Japão conquistou 8 ouros em Athens 2004 e mantém hegemonia consistente desde introdução do judô em 1964, dominando historicamente o esporte de origem japonesa ([NBC Olympics](https://nbc.com), 2024b). A hegemonia francesa em categorias femininas reflete investimento institucional

direcionado, evidenciado por Teddy Riner (França) que acumulou múltiplas medalhas olímpicas, destacando-se como um dos judocas mais condecorados da história.

A detecção de comunidades revelou segregação temporal completa (índice 100.00) em Judô Masculino versus integração estrutural em Judô Feminino (índice 3.12). Este padrão contra-intuitivo captura dinâmica real: categorias femininas introduzidas em 1992 concentram recorrência entre mesmas atletas ao longo de 3-4 ciclos consecutivos, enquanto categorias masculinas fragmentam-se em gerações desconectadas devido a maior profundidade histórica desde 1964.

#### 4.8.6 Boxe: Rivalidade USA-Cuba e Introdução Feminina Recente

A análise identificou USA dominando Boxe Masculino Welterweight e China liderando Boxe Feminino, com distribuição extremamente dispersa masculina (Top 3 = 22.5%) versus concentração feminina extrema (Top 3 = 75.0%). Dados históricos confirmam: USA lidera contagem all-time com 118 medalhas (50 ouros), seguido por Cuba com 80 medalhas (42 ouros) ([Statista, 2024](#)). Hungria contribuiu László Papp (3 ouros 1948-1956), enquanto Cuba produziu Teófilo Stevenson e Félix Savón (3 ouros cada), validando identificação de múltiplas potências regionais.

A densidade feminina  $25.65\times$  superior à masculina reflete introdução recente (2012, apenas 3 edições) versus tradição centenária masculina, comprimindo campo competitivo em janela temporal mínima. Este padrão metodológico — maior densidade em modalidades recentes devido a menor acumulação histórica — emerge consistentemente em todos os esportes femininos analisados.

#### 4.8.7 Síntese da Validação

A confrontação sistemática entre descobertas da análise de redes e estatísticas históricas conhecidas evidencia concordância robusta em três dimensões:

1. **Dominância geográfica:** Rankings de países por métricas de centralidade agregada coincidem com líderes históricos documentados (USA em Basketball/Athletics/Swimming, Brasil em Futebol, Japão em Judô).
2. **Atletas de alta performance:** *PageRank* elevado correlaciona-se com bicampeões e tricampeões olímpicos históricos (Bolt, Lewis, Weissmuller, Fraser, Riner), validando que algoritmo captura hierarquia competitiva efetiva.
3. **Padrões temporais:** Detecção de comunidades identifica corretamente eras de dominância regional conhecidas (Golden Team Hungria 1950s, hegemonia USSR 1970s-1980s, emergência Jamaica 2000s).

Esta validação externa fortalece confiança metodológica na capacidade de métricas de redes complexas capturarem estruturas competitivas reais, para além de artefatos estatísticos ou viés de dados. A convergência entre análise computacional e conhecimento histórico estabelecido indica que modelagem *per-event* preserva fidedignidade a dinâmicas olímpicas autênticas.

## 4.9 Síntese dos Resultados

A análise dos 12 casos de estudo através de modelagem *per-event* revelou:

**Heterogeneidade Estrutural:** Redes variam de 22 a 1.275 atletas, com densidade variando  $5.5\times$ , evidenciando que tipologia esportiva e gênero moldam estruturas topológicas radicalmente distintas.

**Centralidade Normalizada por Percentis:** Ranking baseado em posição relativa intra-rede revela dominância de esportes coletivos (Futebol 45%, Basquete 15%), com distribuição 70% masculino e 30% feminino quando controlado viés de tamanho amostral.

**Padrão Sistemático de Densidade:** Cinco de seis esportes apresentam densidade feminina superior, contradizendo expectativa baseada em longevidade histórica masculina. Este padrão estrutural não implica automaticamente maior centralidade individual cross-network, requerendo normalização para comparações válidas.

**Evolução Temporal Não-Monotônica:** Concentração geográfica reduz até Era Moderna, depois reverte na Era Contemporânea, evidenciando re-centralização competitiva.

**Vulnerabilidade Assimétrica:** Redes exibem robustez a falhas aleatórias mas vulnerabilidade severa a ataques direcionados, característica de topologia *scale-free*.

Estes resultados demonstram que modelagem *per-event* captura dinâmicas competitivas específicas invisíveis em agregações tradicionais por esporte, validando escolha metodológica fundamental e revelando padrões estruturais de relevância teórica para análise de redes competitivas.

## 5 Conclusão

Este trabalho aplicou análise de redes complexas a dados históricos olímpicos através de modelagem *per-event*, revelando padrões estruturais de competição e hierarquias de performance através de métricas de centralidade e detecção de comunidades.

A metodologia desenvolvida integrou quatro componentes principais: (i) construção de 143 redes competitivas direcionadas e ponderadas (uma por evento específico), (ii) seleção de 12 casos de estudo representativos via *Iconic Score*, (iii) aplicação de *PageRank* adaptado ao contexto esportivo, e (iv) caracterização multidimensional de comunidades através do algoritmo de *Louvain*.

Os resultados obtidos confirmam a hipótese central de que análise de redes revela estruturas competitivas latentes não capturadas por métricas tradicionais de contagem de medalhas. Descobertas contra-intuitivas incluem redes femininas sistematicamente mais densas (ratio até  $25.65\times$  em Boxe) e especialização em ouro correlacionada com posições periféricas nas redes.

### 5.1 Síntese dos Achados Principais

A análise dos 12 casos de estudo selecionados (de 143 redes geradas pela modelagem *per-event*), distribuídos em seis modalidades olímpicas ao longo de 120 anos (1896-2016) revelou diferenças estruturais sistemáticas entre tipos de competição esportiva. Esportes coletivos, individuais de performance e individuais de combate produzem redes com propriedades distintivas que refletem características fundamentais de suas dinâmicas competitivas.

**Diferenças Estruturais entre Modalidades:** Natação apresentou maior modularidade (0.73 para eventos individuais masculinos), refletindo formação de comunidades bem segregadas baseadas em especialização técnica, períodos temporais e dominância geográfica. Basquete e Futebol, como esportes coletivos genuínos, produziram redes densamente interconectadas com modularidade próxima a zero (0.003 e 0.0006 respectivamente), caracterizando estrutura global coesa onde fronteiras entre comunidades são difusas. Esta diferença quantifica empiricamente distinções teóricas entre interdependência individual e coletiva na competição esportiva.

**Comunidades Refletem Múltiplos Fatores:** As comunidades detectadas nos 12 casos de estudo não correspondem a agrupamentos simples por período temporal ou nacionalidade, mas emergem de combinações complexas de fatores. A caracterização multidimensional revelou que comunidades em natação apresentam diversidade temporal

moderada (entropia média 1.60), enquanto esportes coletivos exibem diversidade temporal superior (Basquete: 2.46, Futebol: 2.53), refletindo que comunidades em esportes coletivos agregam atletas de múltiplas décadas devido à natureza densamente conectada destas redes. Modalidades de combate individual (Judô, Boxe) demonstraram segregação estrutural ainda mais acentuada (modularidade média 0.81), com fronteiras delimitadas primariamente por categorias de peso. Esta heterogeneidade indica que algoritmos de detecção de comunidades capturam estruturas emergentes não redutíveis a características isoladas.

**Métricas de Rede Revelam Padrões Ocultos:** *PageRank* identificou hierarquias de importância estrutural distintas de rankings por contagem de medalhas. Análise contextualizada por esporte revelou atletas de alta centralidade em cada modalidade: Usain Bolt (Atletismo 100m, *PageRank* 0.0517), Dawn Fraser (Natação 100m Livre, 0.0518), Teresa Edwards (Basquete F, 0.0127), Carmelo Anthony (Basquete M, 0.0057), entre outros. Comparações cross-esporte de valores absolutos são metodologicamente problemáticas devido a diferenças estruturais fundamentais: esportes coletivos (onde equipes recebem medalhas simultaneamente) apresentam topologia radicalmente distinta de esportes individuais (competição isolada). Redes pequenas concentram *PageRank* em poucos nós, enquanto redes massivas diluem centralidade mesmo de campeões históricos. Consequentemente, identificação de atletas centrais foi apresentada de forma independente por modalidade, evitando agregações cross-esporte que violem premissas de comparabilidade. Esta distinção exemplifica como métricas de rede capturam propriedades estruturais invisíveis em análises convencionais, mas exigem interpretação contextual criteriosa respeitando limites metodológicos.

**Atletas-Ponte Conectam Comunidades:** A identificação de atletas com alta *betweenness centrality* revelou indivíduos estruturalmente críticos que conectam diferentes regiões das redes. Em natação, 10.2% dos atletas atuam como pontes, proporção superior aos esportes coletivos (5.3-7.1%). Estes atletas frequentemente apresentam longevidade olímpica ou versatilidade técnica, posicionando-se como conectores entre eras competitivas ou especializações distintas. Sua importância estrutural transcende performance individual, evidenciando papel sistêmico não capturado por métricas tradicionais.

## 5.2 Contribuições Metodológicas

Este trabalho entrega quatro contribuições principais para a literatura de análise de redes aplicada ao esporte:

**Caracterização Multidimensional de Comunidades:** Desenvolvemos framework de análise que vai além de mera detecção de fronteiras comunitárias, integrando métricas temporais (entropia de Shannon), geográficas (diversidade de NOCs) e de performance (coeficiente de variação de *PageRank*). Esta abordagem revelou que comunidades

em esportes individuais exibem alta segregação temporal e geográfica, enquanto esportes coletivos apresentam maior homogeneidade estrutural.

**Modelagem *Per-Event*:** A decisão metodológica de construir redes por evento olímpico específico (Year, Event\_Normalized) ao invés de agregar por esporte demonstrou superioridade analítica ao preservar homogeneidade competitiva intra-rede. Esta abordagem evita mistura artificial de atletas de especializações físico-técnicas distintas (ex: 100m livre vs maratona aquática), gerando 143 redes granulares das quais 12 casos representativos foram selecionados via Iconic Score. A validação através de descobertas contra-intuitivas (densidade feminina sistematicamente superior em 5/6 esportes, especialização em ouro na periferia) evidencia que modelagem *per-event* captura dinâmicas competitivas específicas invisíveis em agregações tradicionais.

**Análise Comparativa Multi-Esporte:** A aplicação da metodologia a seis modalidades olímpicas com diferentes tipos de interdependência competitiva (individual puro, coletivo, misto, combate individual) abrangendo 120 anos de história olímpica permite identificação de padrões estruturais sistemáticos relacionados ao formato competitivo. A diferença de modularidade entre Natação (0.73), esportes de combate (Judô/Boxe: 0.81), e esportes coletivos genuínos (Basquete/Futebol: 0.003) quantifica empiricamente distinções teóricas entre tipos de competição, validando aplicabilidade de análise de redes para caracterização de estruturas competitivas diversas.

**Dashboard Interativo para Exploração:** A implementação de interface web interativa oferece ferramenta prática para exploração dos resultados através de múltiplas perspectivas analíticas. O dashboard integra análise de centralidades, visualização de comunidades, análise temporal e comparações entre esportes, facilitando identificação de padrões e validação de hipóteses sobre estruturas competitivas. Esta contribuição demonstra viabilidade de ferramentas acessíveis para exploração de redes complexas por audiências não-técnicas.

### 5.3 Limitações do Estudo

Este trabalho apresenta limitações que devem ser consideradas na interpretação dos resultados:

**Limitações de Dados:** Registros de atletas pré-1960 são fragmentados, potencialmente sub-representando medalhistas de eras anteriores. Dados incompletos sobre altura, peso e idade limitam análises bioantropométricas. A cobertura temporal até Rio 2016 não inclui edições mais recentes (Tokyo 2020, Paris 2024). O foco em 12 eventos específicos (de 6 esportes) selecionados de 143 redes geradas impossibilita generalização para todas as modalidades olímpicas. A seleção via Iconic Score privilegia eventos de longa trajetória histórica, sub-representando modalidades recentemente introduzidas.

**Limitações Metodológicas:** O sistema de ponderação 5-3-2 carece de fundamentação teórica sólida, sendo validado apenas empiricamente. *PageRank* apresenta diluição estrutural em esportes coletivos devido ao tamanho das redes. O algoritmo de *Louvain* requer simetrização, descartando informação direcional, embora validação com Infomap tenha demonstrado concordância de 85% (NMI). Validação estatística formal através de *null models* não foi implementada, embora diferenças de modularidade (razão de 243×) suportem conclusões qualitativas.

**Limitações de Escopo:** A análise não incorpora dados contextuais (investimento esportivo, PIB, políticas públicas) que impactam estruturas competitivas. O foco exclusivo em medalhistas ignora a estrutura competitiva completa, capturando apenas o topo da hierarquia.

## 5.4 Trabalhos Futuros

Os resultados obtidos e limitações identificadas sugerem múltiplas direções promissoras para extensão desta pesquisa.

**Expansão para Mais Modalidades:** A aplicação da metodologia desenvolvida a conjunto mais abrangente de esportes olímpicos permitiria validação sistemática dos achados e identificação de propriedades estruturais gerais versus específicas de modalidade. Particular interesse reside em esportes de combate com categorias adicionais (taekwondo, luta olímpica), modalidades cronometradas sem confronto direto (ciclismo, remo), eventos artísticos com avaliação subjetiva (ginástica artística, nado sincronizado) e esportes de raquete (tênis, badminton), que apresentam características estruturais potencialmente distintas das seis modalidades analisadas (Natação, Basquete, Futebol, Atletismo, Judô, Boxe).

**Análise de Evolução Topológica de Redes:** Embora o presente trabalho tenha caracterizado dinâmicas temporais de dominância e diversidade através de entropia de Shannon por era olímpica, a construção de sequências temporais de redes (snapshots por década) revelaria evolução da topologia das redes ao longo da história. Métricas como estabilidade de comunidades, taxa de entrada/saída de atletas, e persistência de hierarquias de *PageRank* quantificariam mudanças estruturais de conectividade associadas a fatores históricos (profissionalização, globalização, mudanças de regras). Modelos de redes temporais permitiriam previsão de tendências futuras em estruturas competitivas.

**Incorporação de Atletas Não-Medalhistas:** A expansão da modelagem para incluir todos os competidores (não apenas medalhistas) permitiria caracterizar a estrutura competitiva completa, revelando dinâmicas de eliminação, barreiras de entrada ao pódio e trajetórias de ascensão. Análise de caminhos entre atletas periféricos e centrais identificaria mecanismos estruturais de mobilidade competitiva e fatores que distinguem atletas que

alcançam medalhas daqueles que permanecem em posições marginais.

**Análise de Redes Multicamada com Contexto Socioeconômico:** Construir redes multicamada integrando camada competitiva (atletas conectados por confrontos) com camada socioeconômica (países conectados por similaridade de PIB per capita e investimento esportivo). Hipótese: *PageRank* médio de comunidades nacionais correlaciona positivamente com investimento per capita em esporte ( $r > 0.6$ ), mas saturação ocorre acima de *threshold* (ex: USD 50/habitante/ano), testável via regressão não-linear. Dados da UNESCO e Banco Mundial permitiriam quantificar retorno marginal decrescente de investimento esportivo sobre dominância competitiva.

O presente trabalho estabeleceu fundamentos metodológicos sólidos para aplicação de análise de redes complexas ao contexto olímpico, demonstrando viabilidade e valor analítico da abordagem. As direções futuras propostas expandiriam compreensão de estruturas competitivas esportivas, contribuindo tanto para avanço teórico em ciência de redes quanto para aplicações práticas em gestão esportiva e políticas de desenvolvimento atlético.

# Referências

- AHMED, A. et al. Research progress of complex network modeling methods based on uncertainty theory. *Mathematics*, v. 11, n. 5, p. 1212, 2023. Citado 2 vezes nas páginas 17 e 21.
- Alpha History. *Sport in the Cold War*. 2024. Alpha History. Análise histórica da rivalidade esportiva USA-USSR durante Guerra Fria (1952-1991). Sucesso olímpico era visto como indicador de superioridade ideológica. Inclui confrontos emblemáticos como basquete 1972 e hockey "Miracle on Ice"1980. Disponível em: <<https://alphahistory.com/coldwar/sport-cold-war/>>. Citado na página 89.
- BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. *Science*, v. 286, n. 5439, p. 509–512, 1999. Citado na página 22.
- BLONDEL, V. D. et al. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, v. 2008, n. 10, p. P10008, 2008. Citado 3 vezes nas páginas 27, 46 e 48.
- BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. In: *Computer Networks and ISDN Systems*. [S.l.: s.n.], 1998. v. 30, n. 1-7, p. 107–117. Citado na página 25.
- DAVIDS, K. et al. *Complex Systems in Sport*. [S.l.]: Routledge, 2013. ISBN 9780415809702. Citado na página 17.
- ERIKSSON, A. et al. Mapping nonlocal relationships between metadata and network structure with metadata-dependent encoding of random walks. *Science Advances*, v. 8, n. 30, p. eabn7558, 2022. Citado 2 vezes nas páginas 28 e 29.
- FABBRI, F. et al. Inequality and inequity in network-based ranking and recommendation algorithms. *Scientific Reports*, v. 12, p. 2012, 2022. Citado na página 29.
- FARIA, J. G. R. de; FERREIRA, F. G. Podium and influence: A network analysis of the most important formula one drivers. In: *Anais do XIII Brazilian Workshop on Social Network Analysis and Mining*. Porto Alegre, RS, Brasil: SBC, 2024. p. 144–157. ISSN 2595-6094. Disponível em: <<https://sol.sbc.org.br/index.php/brasnam/article/view/29339>>. Citado 3 vezes nas páginas 25, 31 e 34.
- FORTUNATO, S. Community detection in graphs. *Physics Reports*, v. 486, n. 3-5, p. 75–174, 2010. Citado 4 vezes nas páginas 22, 26, 27 e 28.
- FREEMAN, L. C. Centrality in social networks: Conceptual clarification. *Social Networks*, v. 1, n. 3, p. 215–239, 1979. Citado 2 vezes nas páginas 24 e 25.
- GRIFFIN, R. *120 years of Olympic history: athletes and results*. 2018. Kaggle. Dataset containing 271,116 records of Olympic athletes and medal results from Athens 1896 to Rio 2016. Disponível em: <<https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results>>. Citado 2 vezes nas páginas 39 e 40.

HAGBERG, A. A.; SCHULT, D. A.; SWART, P. J. Exploring network structure, dynamics, and function using NetworkX. In: VAROQUAUX, G.; VAUGHT, T.; MILLMAN, J. (Ed.). *Proceedings of the 7th Python in Science Conference (SciPy2008)*. Pasadena, CA USA: [s.n.], 2008. p. 11–15. Citado na página 36.

HSU, Y.-C.; ZHANG, W.-J. Applying pagerank to team ranking in single-elimination tournaments: Evidence from taiwan's high school baseball. *Applied Sciences*, v. 15, n. 12, p. 6882, 2025. Citado 6 vezes nas páginas 17, 24, 25, 26, 45 e 60.

JACOMY, M. et al. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS ONE*, v. 9, n. 6, p. e98679, 2014. Citado na página 53.

LEICHT, E. A.; NEWMAN, M. E. J. Community structure in directed networks. *Physical Review Letters*, v. 100, n. 11, p. 118703, 2008. Citado 4 vezes nas páginas 21, 27, 28 e 46.

LÓPEZ-FELIP, M. A. et al. A cluster phase analysis for collective behavior in team sports. *Human Movement Science*, v. 59, p. 96–111, 2018. Citado na página 17.

MCPHERSON, M.; SMITH-LOVIN, L.; COOK, J. M. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, v. 27, p. 415–444, 2001. Citado na página 22.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to Linear Regression Analysis*. 6th. ed. Hoboken, NJ: Wiley, 2021. Citado na página 31.

MOTEGI, S.; MASUDA, N. A network-based dynamical ranking system for competitive sports. *Scientific Reports*, v. 2, p. 904, 2012. Citado 4 vezes nas páginas 17, 23, 26 e 44.

NBC Olympics. *Basketball 101: Olympic history, records and results*. 2024. NBC Olympics. Histórico completo do basquete olímpico. USA conquistou 16 de 20 torneios masculinos (80%) e 9 de 12 torneios femininos, estabelecendo domínio histórico. União Soviética (4 ouros) em segundo lugar. Inclui confronto controverso USA-USSR de 1972 (51-50). Disponível em: <<https://www.nbcolympics.com/news/basketball-101-olympic-history-records-and-results>>. Citado na página 89.

NBC Olympics. *Judo 101: Olympic history, records and results*. 2024. NBC Olympics. Histórico do judô olímpico desde 1964. Japão domina historicamente: 9 ouros em Tokyo 2021 (recorde olímpico) e 8 ouros em Athens 2004. Teddy Riner (França) detém recorde individual com 6 medalhas (4 ouros). Nomura (Japão) único tricampeão. Disponível em: <<https://www.nbcolympics.com/news/judo-101-olympic-history-records-and-results>>. Citado na página 90.

NEWMAN, M. E. J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, v. 103, n. 23, p. 8577–8582, 2006. Citado na página 57.

NEWMAN, M. E. J. *Networks*. 2. ed. Oxford, UK: Oxford University Press, 2018. ISBN 978-0-19-880509-0. Citado na página 30.

Olympics.com. *The legend of the Hungarian Golden Team was born at the Helsinki Games in 1952*. 2024. Website Oficial dos Jogos Olímpicos. Documentação histórica do time húngaro de futebol que revolucionou o esporte nos anos 1950. Entre 1950-1956, o time registrou 58 vitórias, 10 empates e apenas 1 derrota, conquistando ouro em

Helsinki 1952 com Ferenc Puskás. Disponível em: <<https://www.olympics.com/en/news/the-legend-of-the-hungarian-golden-team-was-born-at-the-helsinki-games-in-1952>>. Citado na página 89.

Olympics.com. *Olympic football winners list: Know the champions*. 2024. Website Oficial dos Jogos Olímpicos. Lista completa de campeões olímpicos de futebol masculino e feminino desde 1900. Brasil lidera contagem total de medalhas masculinas com 6 medalhas (1 ouro, 3 prata, 2 bronze até 2016). Hungria e Grã-Bretanha empatadas com 3 ouros cada. Disponível em: <<https://www.olympics.com/en/news/olympic-football-winners-list-men-women-gold-medals-champions>>. Citado na página 89.

Olympics.com. *Olympics 100m winners list: Know the champions*. 2024. Website Oficial dos Jogos Olímpicos. Lista completa de campeões olímpicos dos 100m rasos desde 1896. USA conquistou 16 de 29 ouros masculinos (55%) e 9 de 22 ouros femininos. Usain Bolt (Jamaica) único tricampeão olímpico. Carl Lewis (USA) bicampeão. Disponível em: <<https://www.olympics.com/en/news/olympics-100-metres-winners-list-men-women-gold-medals-champions>>. Citado na página 90.

PALINKAS, L. A. et al. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, v. 42, n. 5, p. 533–544, 2015. Citado 2 vezes nas páginas 17 e 42.

PEREIRA-FERRERO, V. H. et al. Complex networks models and spectral decomposition in the analysis of swimming athletes' performance at olympic games. *Frontiers in Physiology*, v. 10, p. 1134, 2019. Citado 2 vezes nas páginas 32 e 34.

PEREIRA, V. H. et al. Computational and complex network modeling for analysis of sprinter athletes' performance in track field tests. *Frontiers in Physiology*, v. 9, p. 843, 2018. Citado na página 41.

RADICCHI, F. Analysing olympic games through dominance networks. *Physica A: Statistical Mechanics and its Applications*, v. 462, p. 1215–1227, 2016. Citado 3 vezes nas páginas 17, 23 e 42.

RAJARAM, R.; RITCHEY, N. P.; CASTELLANI, B. Advancing shannon entropy for measuring diversity in systems. *Complexity*, v. 2017, p. 8715605, 2017. Citado na página 48.

ROSVALL, M.; BERGSTROM, C. T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, v. 105, n. 4, p. 1118–1123, 2008. Citado 2 vezes nas páginas 27 e 46.

Statista. *Olympic boxing all time medal table 2022*. 2024. Statista. Tabela histórica de medalhas do boxe olímpico. USA lidera com 118 medalhas (50 ouros), seguido por Cuba com 80 medalhas (42 ouros). Cuba emergiu como potência no período 1972-2000 com três lutadores tricampeões: Teófilo Stevenson e Félix Savón. Disponível em: <<https://www.statista.com/statistics/1339565/country-most-olympic-medals-boxing/>>. Citado na página 91.

TopEndSports. *100m Freestyle Swimming at the Olympic Games*. 2024. TopEndSports. Histórico do evento de natação 100m nado livre desde 1896. Duke Kahanamoku, Johnny Weissmuller (USA), Alexander Popov (RUS), e Pieter van den Hoogenband (NED) são os únicos bicampeões. Dawn Fraser (AUS) tricampeã feminina 1956-1964. Disponível em: <<https://www.topendsports.com/events/summer/sports/swimming-100m-freestyle.htm>>. Citado na página 90.

VANDERPLAS, J. *Python Data Science Handbook: Essential Tools for Working with Data*. Sebastopol, CA: O'Reilly Media, 2016. ISBN 978-1-491-91205-8. Citado 3 vezes nas páginas 30, 36 e 63.

VEGA-OLIVEROS, D.; ZHAO, L.; BERTON, L. Complex networks for community detection of basketball players. *Annals of Operations Research*, v. 325, p. 363–400, 2023. Citado 2 vezes nas páginas 33 e 34.

VINH, N. X.; EPPS, J.; BAILEY, J. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, v. 11, p. 2837–2854, 2010. Citado na página 56.

WÄSCHE, H. et al. Social network analysis in sport research: an emerging paradigm. *European Journal for Sport and Society*, v. 14, n. 2, p. 138–165, 2017. Citado na página 17.