CIÊNCIA DE DADOS (BIG DATA PROCESSING AND ANALYTICS)

RECUPERAÇÃO DA INFORMAÇÃO NA WEB E EM REDES SOCIAIS

Professor curador
Luciano Moreira Camilo e Silva

TRILHA 1
INTRODUÇÃO À RECUPERAÇÃO DA INFORMAÇÃO

Recuperação de dados *versus* Recuperação da informação

Recuperação de dados

determinístico • estruturado • exato

Fonte: GettyImages.

Recuperação da informação

probabilístico • não-estruturado • relevante

Fonte: GettyImages.

Recuperação de dados *versus* Recuperação da informação

Fonte: Elaborada pelo autor.

Sistemas de Recuperação da informação

- Disciplina da grande área das ciências sociais (biblioteconomia).
- Evolução em parceria com a disciplina de sistemas da informação.
- Tornou-se dominante após a década de 1990, como a ferramenta *de facto* de recuperação de documentos na web.

Sistema ZATOR 800 de recuperação da informação de 1951.
Fonte: MOOERS, C. N. *Aslib Proceedings*, v. 8, n. 1, p. 3-22, [s.d.].

Modelos de sistema de Recuperação da informação

Modelo Booleano

São Paulo

Santo

Cidade

Modelo Vetorial

Paulo

Consulta:
"São Paulo" E "Cidade"

7

θ_7

9

θ_9

θ_8

8

Cidade

Fonte: Elaborada pelo autor.

Índices Invertidos

No meio do caminho tinha uma pedra
Tinha uma pedra no meio do caminho
Tinha uma pedra
No meio do caminho tinha uma pedra
Nunca me esquecerei desse acontecimento
Na vida de minhas retinas tão fatigadas
Nunca me esquecerei que no meio do caminho
Tinha uma pedra
Tinha uma pedra no meio do caminho
No meio do caminho tinha uma pedra.

Carlos Drummond de Andrade

Índices Invertidos

1. *No meio do caminho tinha uma pedra*
2. *Tinha uma pedra no meio do caminho*
3. *Tinha uma pedra*
4. *No meio do caminho tinha uma pedra*
5. *Nunca me esquecerei desse acontecimento*
6. *Na vida de minhas retinas tão fatigadas*
7. *Nunca me esquecerei que no meio do caminho*
8. *Tinha uma pedra*
9. *Tinha uma pedra no meio do caminho*
10. *No meio do caminho tinha uma pedra.*

Índices Invertidos

1. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra***
2. ***tinha*** *uma* ***pedra*** *no* ***meio*** *do* ***caminho***
3. ***tinha*** *uma* ***pedra***
4. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra***
5. ***nunca*** *me* ***esquecerei*** *desse* ***acontecimento***
6. *na* ***vida*** *de minhas* ***retinas*** *tão* ***fatigadas***
7. ***nunca*** *me* ***esquecerei*** *que no* ***meio*** *do* ***caminho***
8. ***tinha*** *uma* ***pedra***
9. ***tinha*** *uma* ***pedra*** *no* ***meio*** *do* ***caminho***
10. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra.***

Índices Invertidos

1. no meio do caminho tinha uma pedra
2. tinha uma pedra no meio do caminho
3. tinha uma pedra
4. no meio do caminho tinha uma pedra
5. nunca me esquecerei desse acontecimento
6. na vida de minhas retinas tão fatigadas
7. nunca me esquecerei que no meio do caminho
8. tinha uma pedra
9. tinha uma pedra no meio do caminho
10. no meio do caminho tinha uma pedra.

Word	Doc_Ids
	5
	1 2 4 7 9 10

Índices Invertidos

1. *no meio do caminho tinha uma pedra*
2. *tinha uma pedra no meio do caminho*
3. *tinha uma pedra*
4. *no meio do caminho tinha uma pedra*
5. *nunca me esquecerei desse acontecimento*
6. *na vida de minhas retinas tão fatigadas*
7. *nunca me esquecerei que no meio do caminho*
8. *tinha uma pedra*
9. *tinha uma pedra no meio do caminho*
10. *no meio do caminho tinha uma pedra.*

Word	Doc_Ids
acontecimento	5
caminho	1,2,4,7,9,10

Índices Invertidos

1. *no meio do caminho tinha uma pedra*
2. *tinha uma pedra no meio do caminho*
3. *tinha uma pedra*
4. *no meio do caminho tinha uma pedra*
5. *nunca me esquecerei desse acontecimento*
6. *na vida de minhas retinas tão fatigadas*
7. *nunca me esquecerei que no meio do caminho*
8. *tinha uma pedra*
9. *tinha uma pedra no meio do caminho*
10. *no meio do caminho tinha uma pedra.*

Word	Doc_Ids
acontecimento	5
caminho	1,2,4,7,9,10
esquecerei	5,7

Índices Invertidos

1. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra***
2. ***tinha*** *uma* ***pedra*** *no* ***meio*** *do* ***caminho***
3. ***tinha*** *uma* ***pedra***
4. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra***
5. ***nunca*** *me* ***esquecerei*** *desse* ***acontecimento***
6. *na* ***vida*** *de minhas* ***retinas*** *tão* ***fatigadas***
7. ***nunca*** *me* ***esquecerei*** *que no* ***meio*** *do* ***caminho***
8. ***tinha*** *uma* ***pedra***
9. ***tinha*** *uma* ***pedra*** *no* ***meio*** *do* ***caminho***
10. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra.***

Word	Doc_Ids
acontecimento	5
caminho	1,2,4,7,9,10
esquecerei	5,7
fatigados	6

Índices Invertidos

1. *no meio do caminho tinha uma pedra*
2. *tinha uma pedra no meio do caminho*
3. *tinha uma pedra*
4. *no meio do caminho tinha uma pedra*
5. *nunca me esquecerei desse acontecimento*
6. *na vida de minhas retinas tão fatigadas*
7. *nunca me esquecerei que no meio do caminho*
8. *tinha uma pedra*
9. *tinha uma pedra no meio do caminho*
10. *no meio do caminho tinha uma pedra.*

Word	Doc_Ids
acontecimento	5
caminho	1,2,4,7,9,10
esquecerei	5,7
fatigados	6
meio	1,2,4,7,9,10

Índices Invertidos

1. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra***
2. ***tinha*** *uma* ***pedra*** *no* ***meio*** *do* ***caminho***
3. ***tinha*** *uma* ***pedra***
4. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra***
5. ***nunca*** *me* ***esquecerei*** *desse* ***acontecimento***
6. *na* ***vida*** *de minhas* ***retinas*** *tão* ***fatigadas***
7. ***nunca*** *me* ***esquecerei*** *que no* ***meio*** *do* ***caminho***
8. ***tinha*** *uma* ***pedra***
9. ***tinha*** *uma* ***pedra*** *no* ***meio*** *do* ***caminho***
10. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra.***

Word	Doc_Ids
acontecimento	5
caminho	1,2,4,7,9,10
esquecerei	5,7
fatigados	6
meio	1,2,4,7,9,10
nunca	5,7

Índices Invertidos

1. *no meio do caminho tinha uma pedra*
2. *tinha uma pedra no meio do caminho*
3. *tinha uma pedra*
4. *no meio do caminho tinha uma pedra*
5. *nunca me esquecerei desse acontecimento*
6. *na vida de minhas retinas tão fatigadas*
7. *nunca me esquecerei que no meio do caminho*
8. *tinha uma pedra*
9. *tinha uma pedra no meio do caminho*
10. *no meio do caminho tinha uma pedra.*

Word	Doc_Ids
acontecimento	5
caminho	1,2,4,7,9,10
esquecerei	5,7
fatigados	6
meio	1,2,4,7,9,10
nunca	5,7
pedra	1,2,3,4,8,9,10

Índices Invertidos

1. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra***
2. ***tinha*** *uma* ***pedra*** *no* ***meio*** *do* ***caminho***
3. ***tinha*** *uma* ***pedra***
4. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra***
5. ***nunca*** *me* ***esquecerei*** *desse* ***acontecimento***
6. *na* ***vida*** *de minhas* ***retinas*** *tão* ***fatigadas***
7. ***nunca*** *me* ***esquecerei*** *que no* ***meio*** *do* ***caminho***
8. ***tinha*** *uma* ***pedra***
9. ***tinha*** *uma* ***pedra*** *no* ***meio*** *do* ***caminho***
10. *no* ***meio*** *do* ***caminho tinha*** *uma* ***pedra.***

Word	Doc_Ids
acontecimento	5
caminho	1,2,4,7,9,10
esquecerei	5,7
fatigados	6
meio	1,2,4,7,9,10
nunca	5,7
pedra	1,2,3,4,8,9,10
retinas	6

Índices Invertidos

1. no meio do caminho tinha uma pedra
2. tinha uma pedra no meio do caminho
3. tinha uma pedra
4. no meio do caminho tinha uma pedra
5. nunca me esquecerei desse acontecimento
6. na vida de minhas retinas tão fatigadas
7. nunca me esquecerei que no meio do caminho
8. tinha uma pedra
9. tinha uma pedra no meio do caminho
10. no meio do caminho tinha uma pedra.

Word	Doc_Ids
acontecimento	5
caminho	1,2,4,7,9,10
esquecerei	5,7
fatigados	6
meio	1,2,4,7,9,10
nunca	5,7
pedra	1,2,3,4,8,9,10
retinas	6
tinha	1,2,3,4,8,9,10

Índices Invertidos

1. no meio do caminho tinha uma pedra
2. tinha uma pedra no meio do caminho
3. tinha uma pedra
4. no meio do caminho tinha uma pedra
5. nunca me esquecerei desse acontecimento
6. na vida de minhas retinas tão fatigadas
7. nunca me esquecerei que no meio do caminho
8. tinha uma pedra
9. tinha uma pedra no meio do caminho
10. no meio do caminho tinha uma pedra.

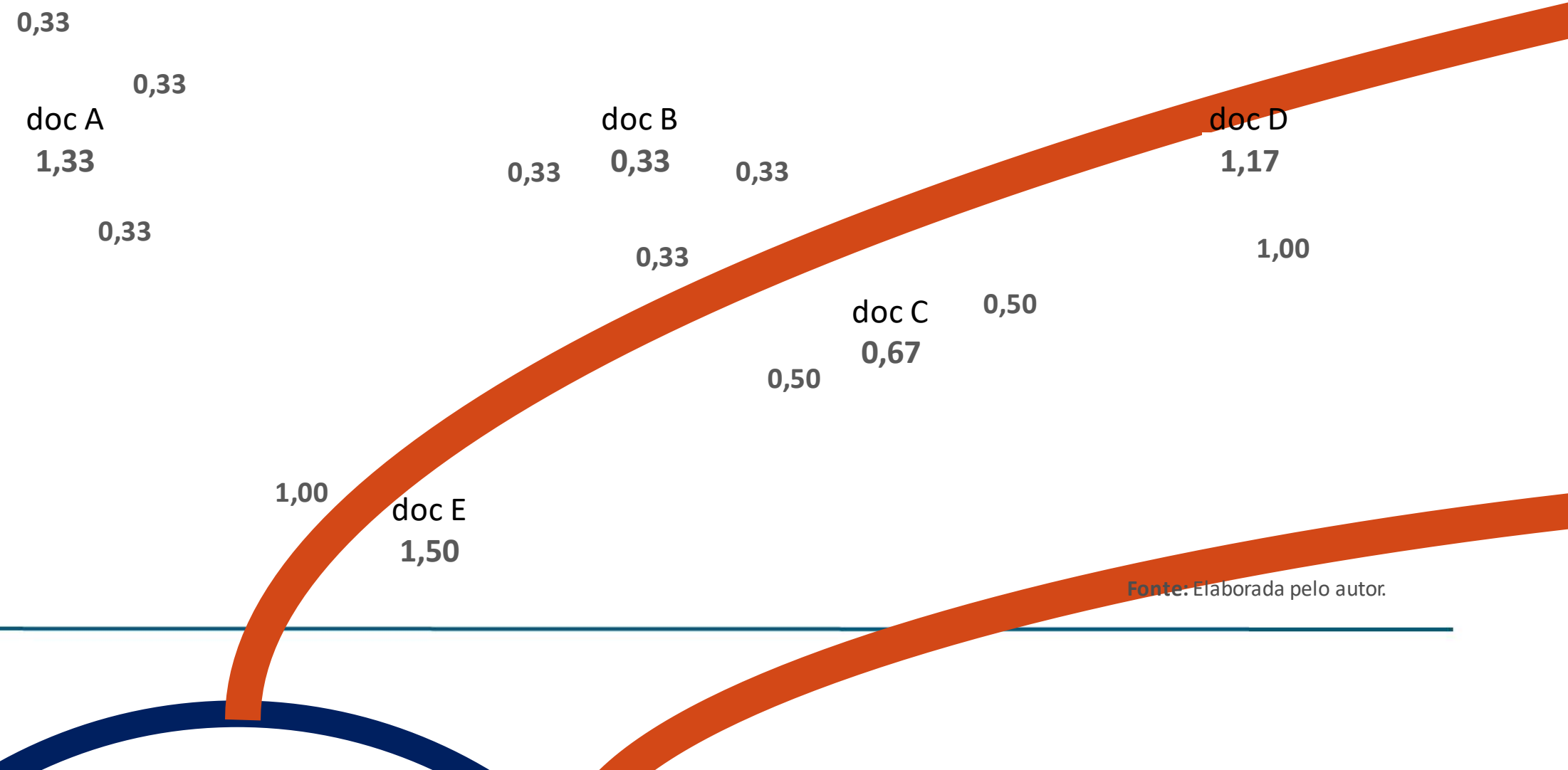
Word	Doc_Ids
acontecimento	5
caminho	1,2,4,7,9,10
esquecerei	5,7
fatigados	6
meio	1,2,4,7,9,10
nunca	5,7
pedra	1,2,3,4,8,9,10
retinas	6
tinha	1,2,3,4,8,9,10
vida	6

Algoritmo PageRank

- Classificação das páginas da web baseada em reputação.
- Cada documento recebe e doa pontos baseados em hiperlinks e em sua reputação.
- Criado pelo cofundador do Google, Larry Page.

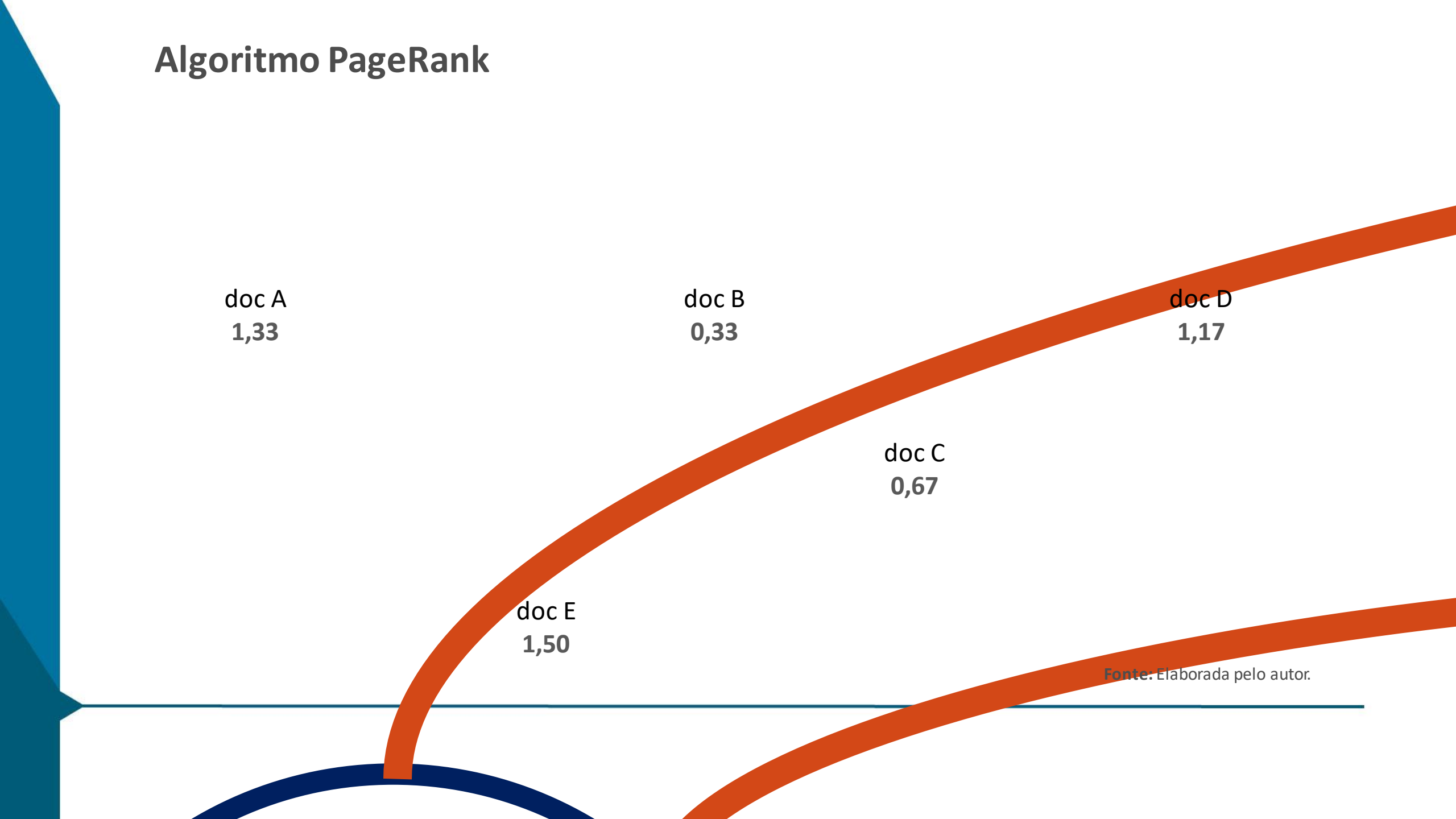
Fonte: Wikimedia.

Algoritmo PageRank



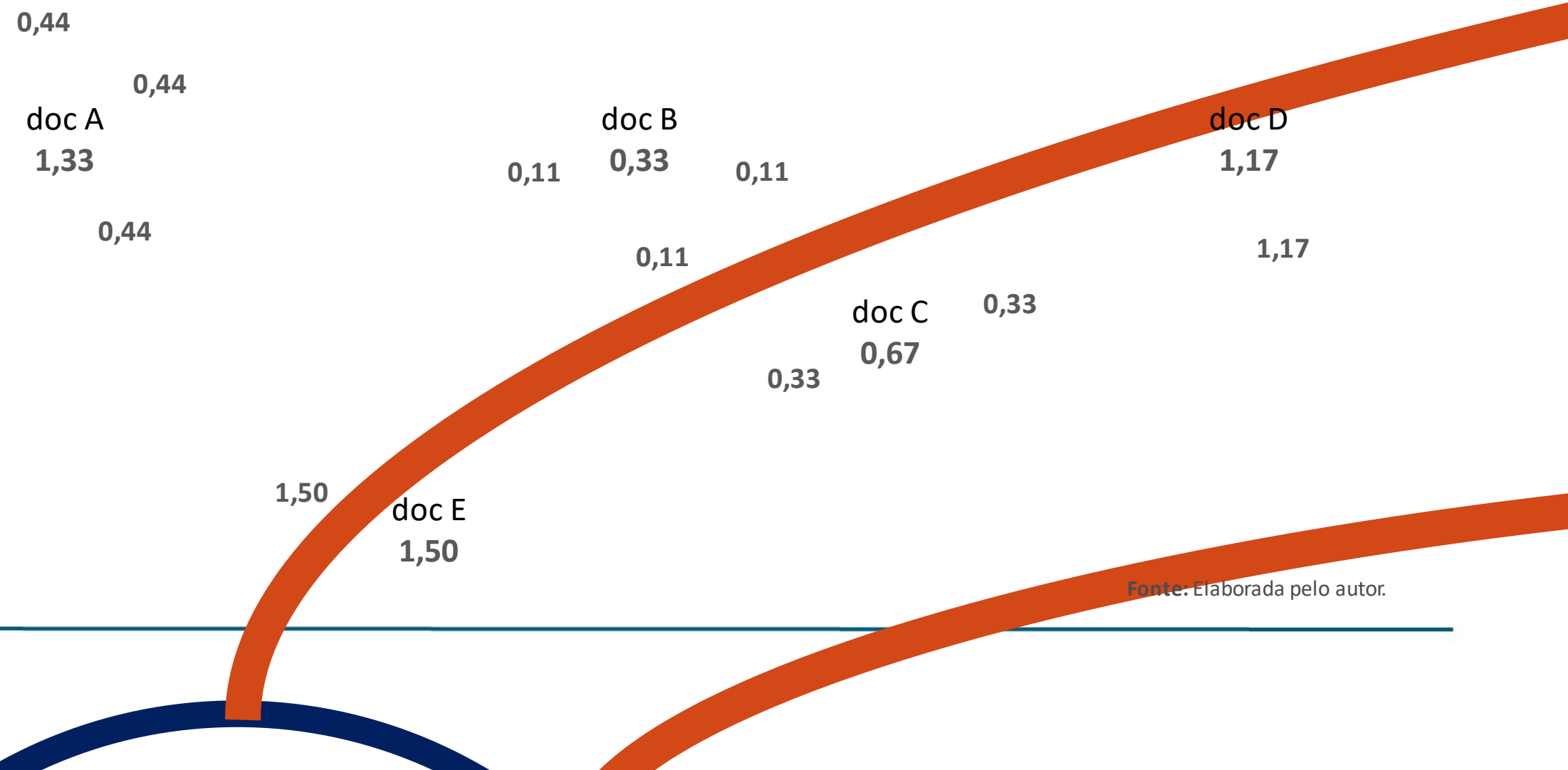
Fonte: Elaborada pelo autor.

Algoritmo PageRank



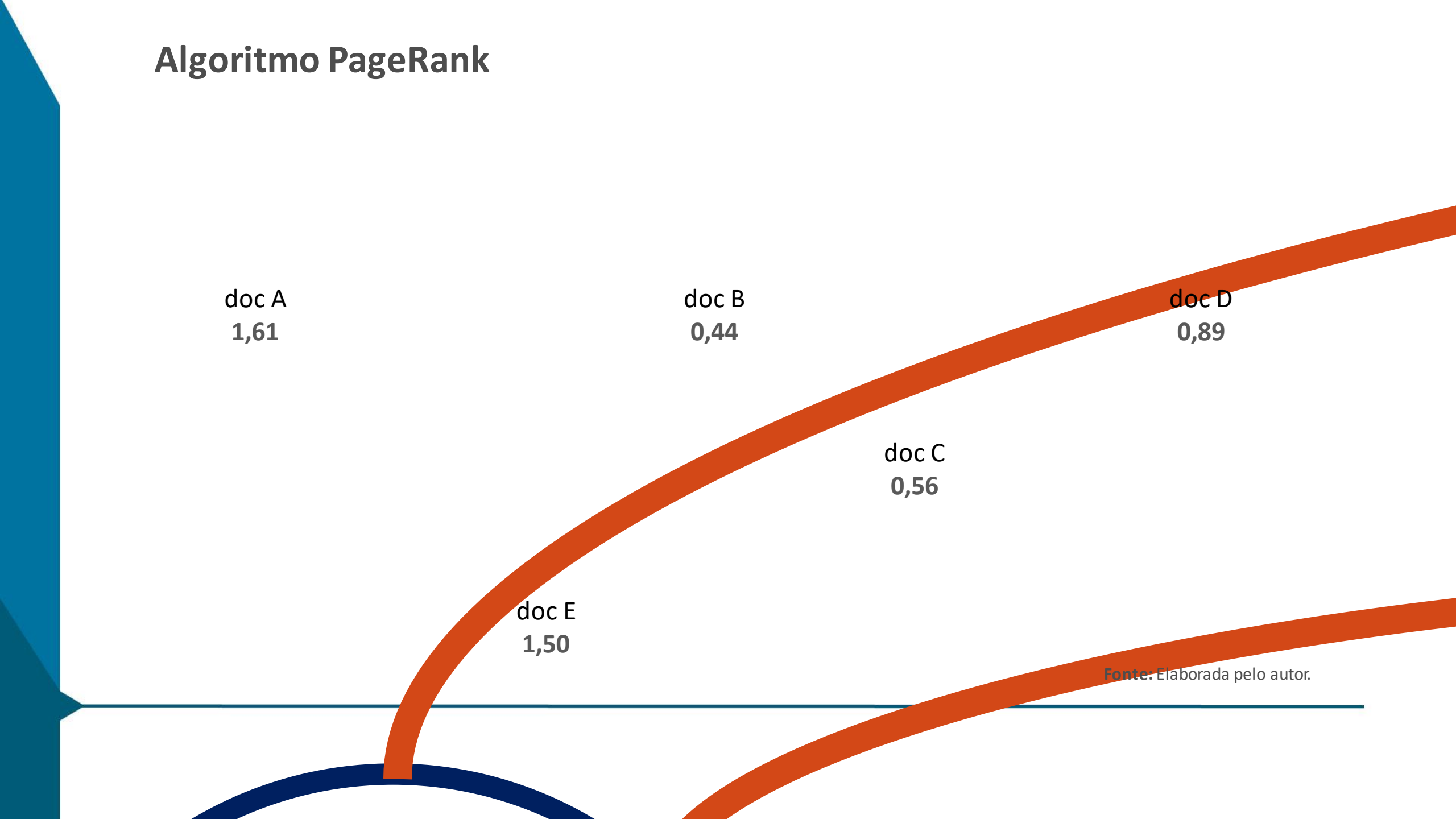
Fonte: Elaborada pelo autor.

Algoritmo PageRank



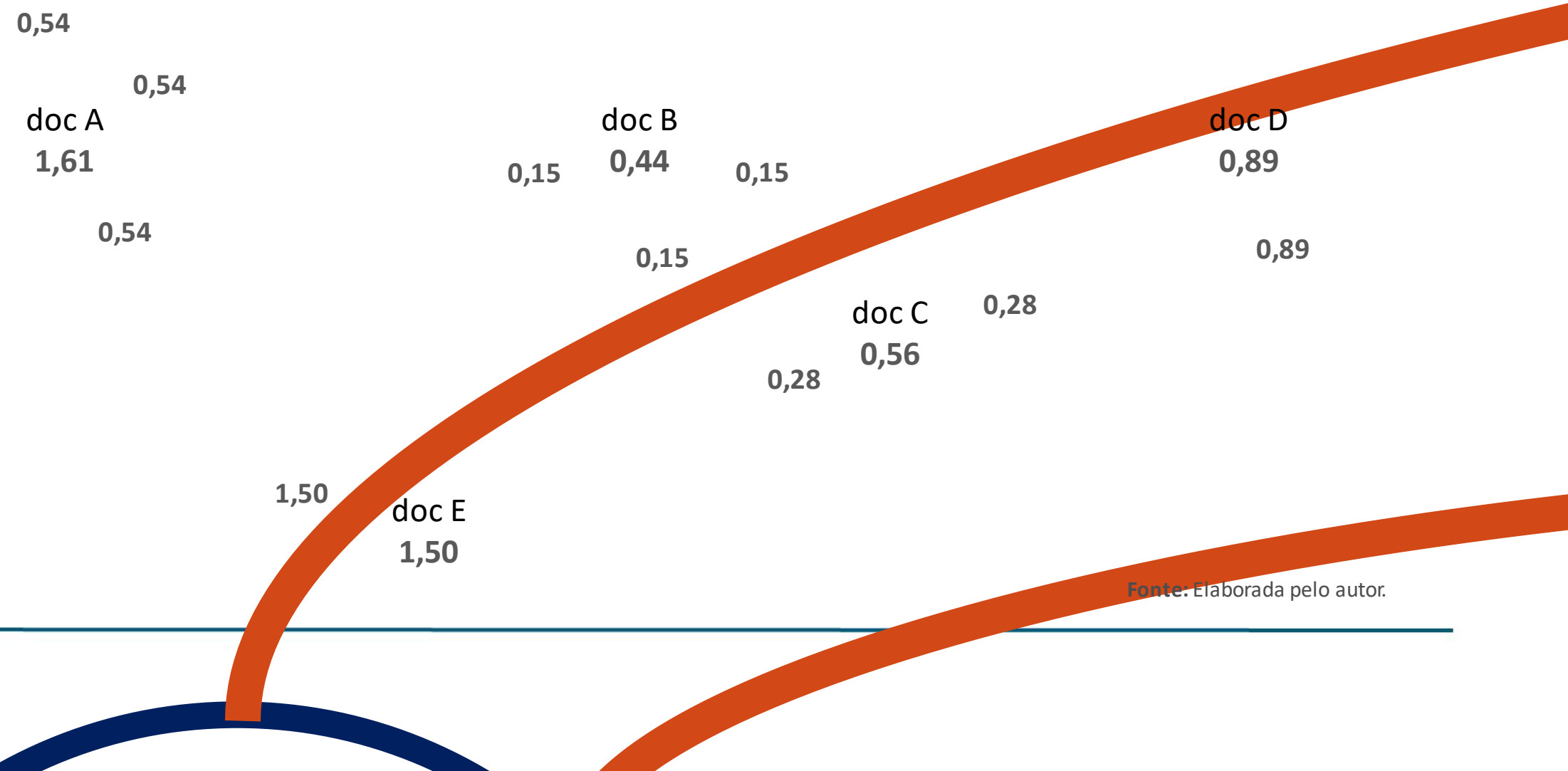
Fonte: Elaborada pelo autor.

Algoritmo PageRank



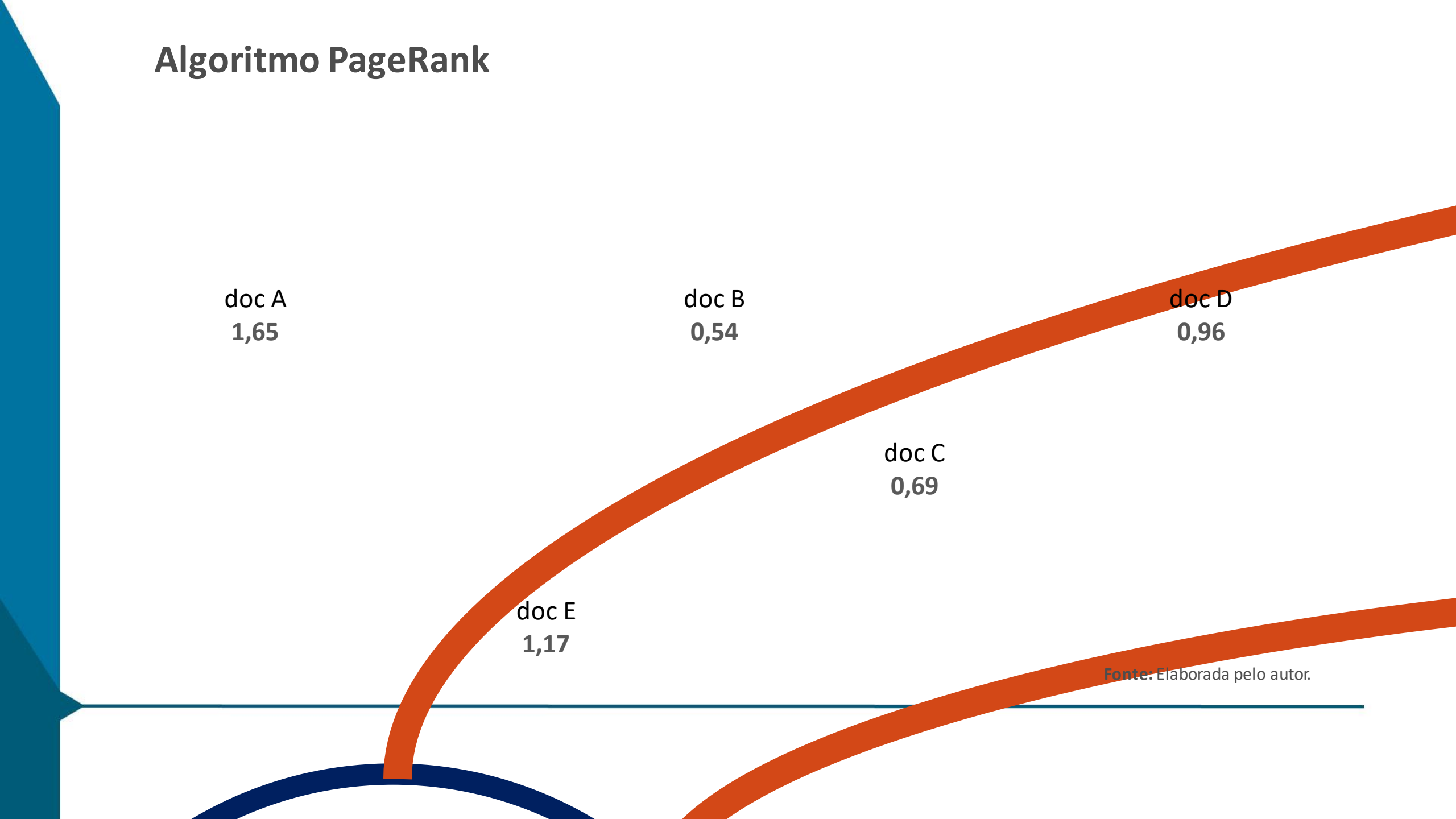
Fonte: Elaborada pelo autor.

Algoritmo PageRank



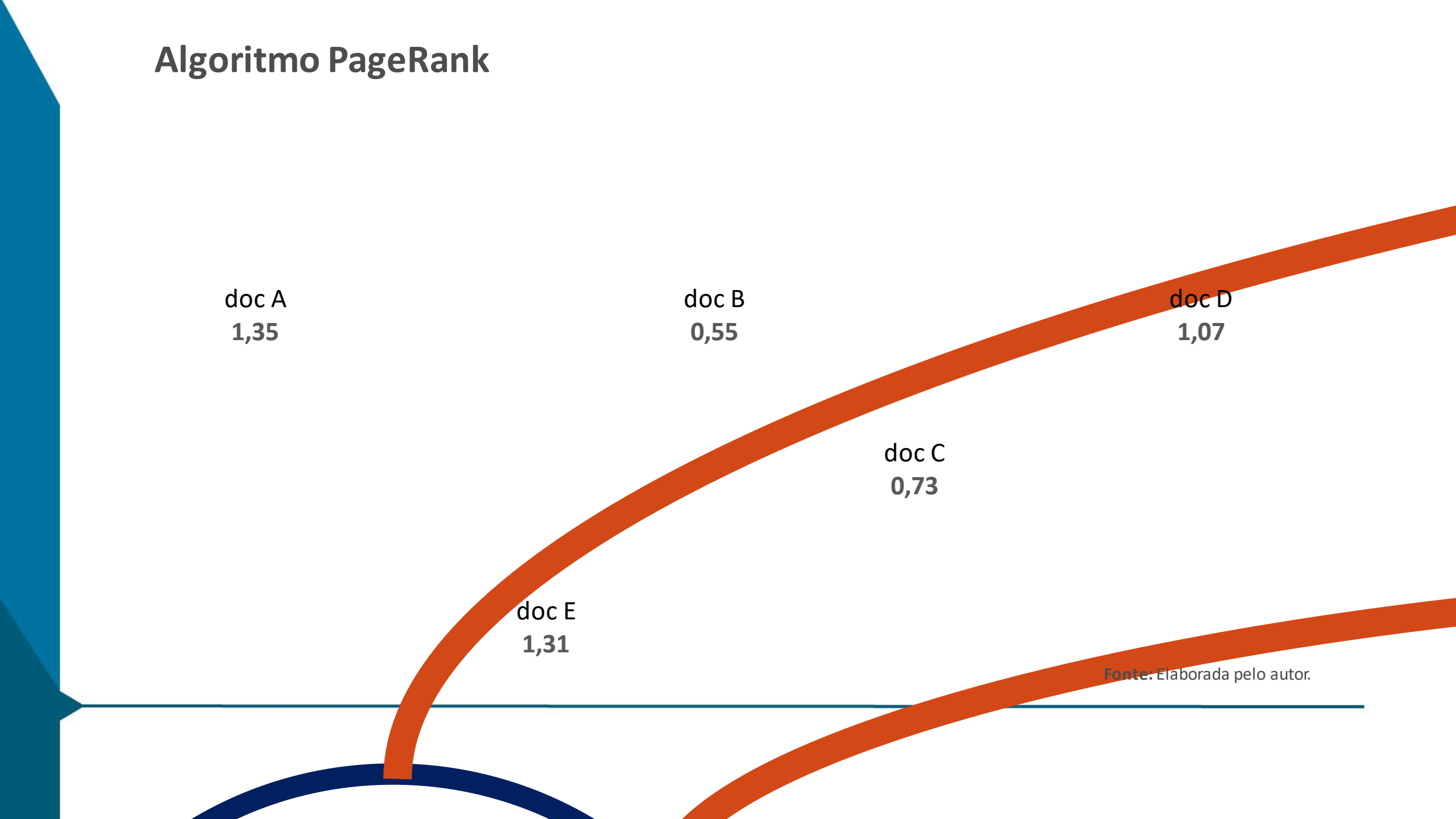
Fonte: Elaborada pelo autor.

Algoritmo PageRank



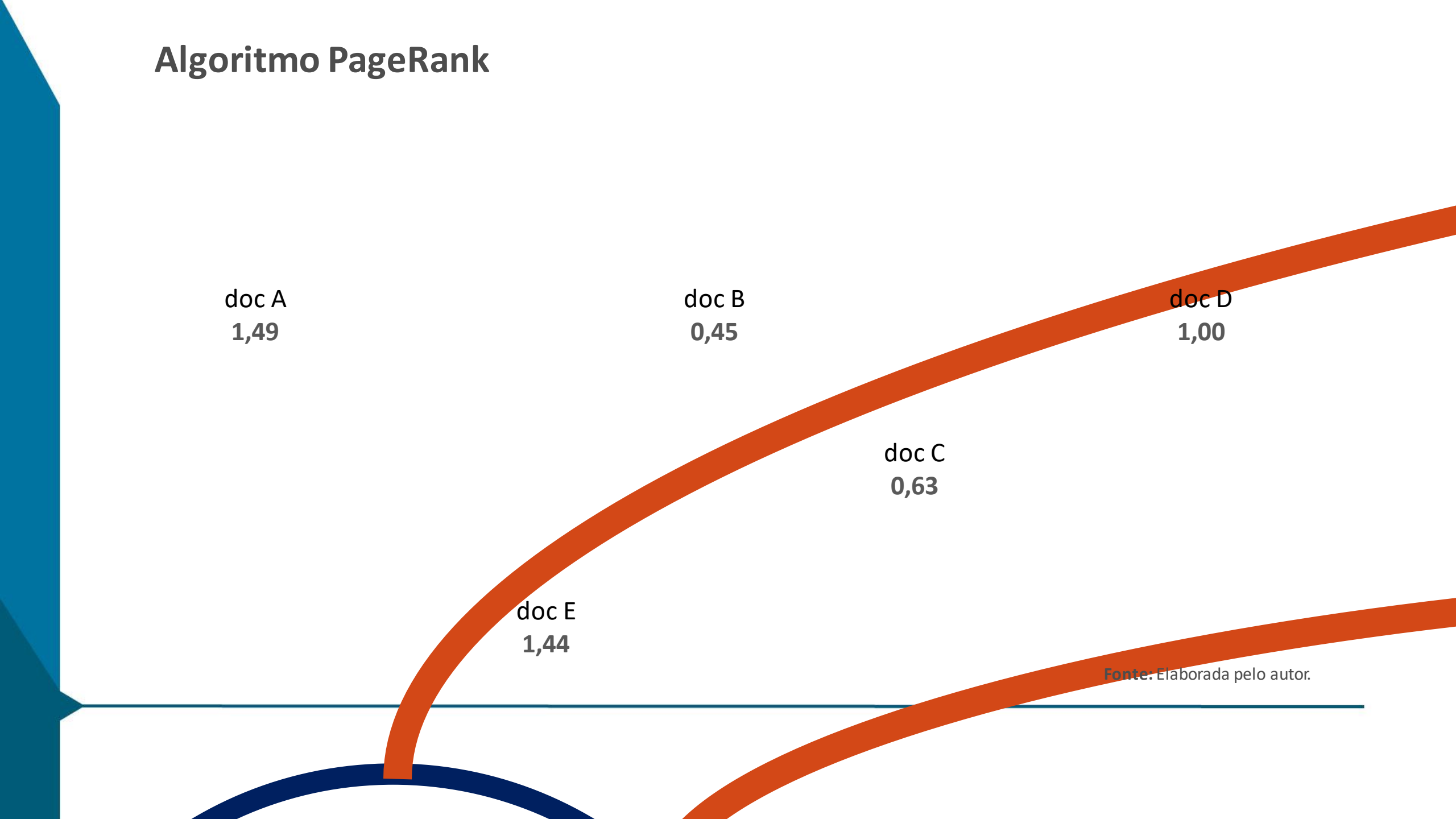
Fonte: Elaborada pelo autor.

Algoritmo PageRank



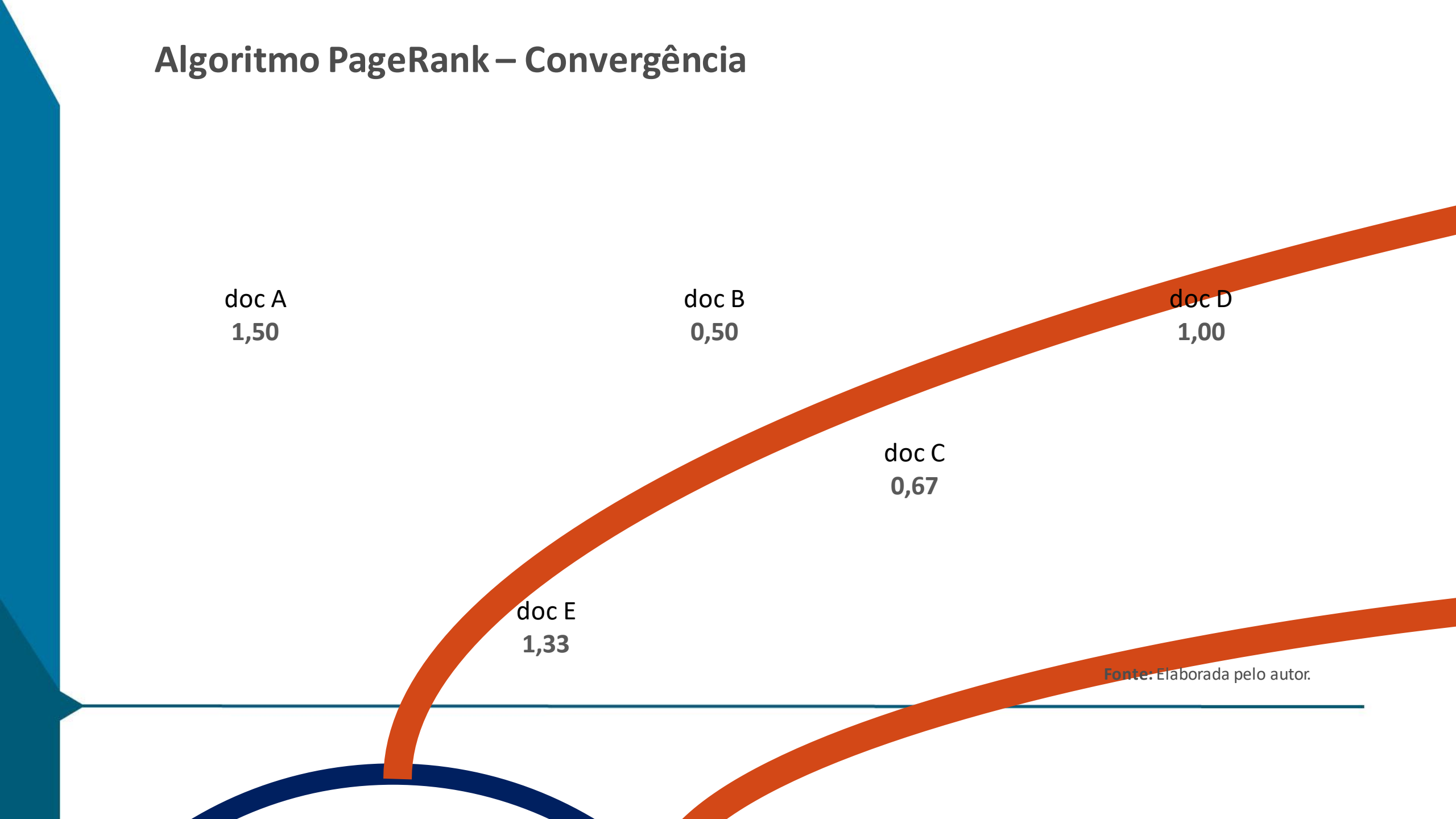
Fonte: Elaborada pelo autor.

Algoritmo PageRank



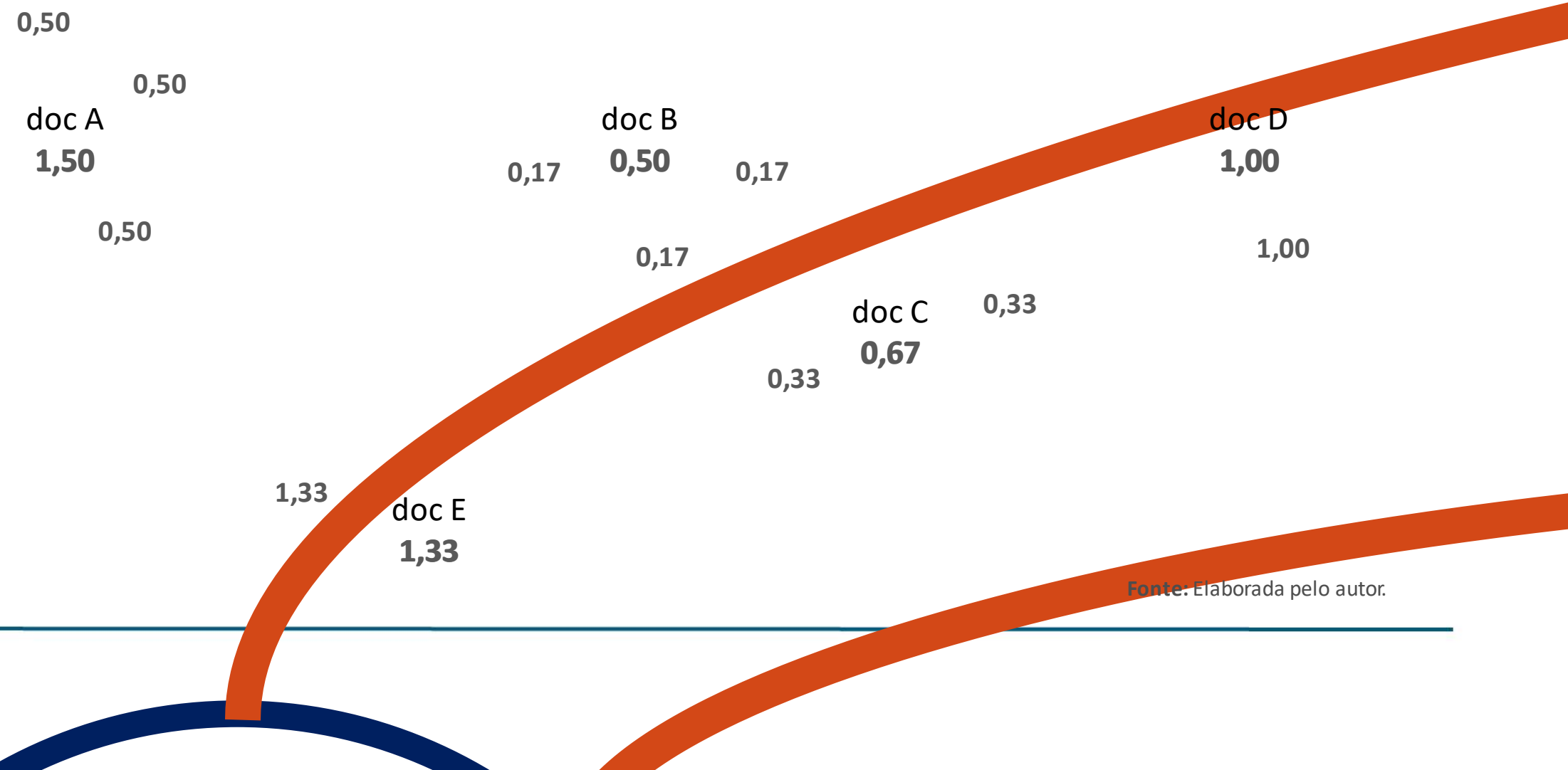
Fonte: Elaborada pelo autor.

Algoritmo PageRank – Convergência



Fonte: Elaborada pelo autor.

Algoritmo PageRank – Convergência



Fonte: Elaborada pelo autor.

Operadores avançados em sistemas de Recuperação da informação

Operador	Para que serve	Exemplo
-palavra	Exclui os resultados que possuírem esta palavra.	“ontologia”-filosofia
“palavra”	Busca pela frase exata e não considera sinônimos ou conjunções de gênero, número e grau.	“worked at night”
“dentro de uma frase”	Busca por uma frase exata, e não pela combinação de palavras.	“melhor carro da formula 1”
..	Procura considerando valores numéricos dentro do intervalo de números.	presidente 2002**..**2020
intitle:	Pesquisa apenas no título da página do documento.	Election **intitle**:biden
inrul:	Pesquisa apenas na URL de identificação do documento.	eleição **inurl**:2018
filetype:	Especifica o formato de documentos procurado (por exemplo: PDF, PPT, DOC, XLS).	**filetype**: pdf
site:	Faz a pesquisa exclusivamente dentro do site especificado.	**site:**senado.gov.br

Universidade Presbiteriana
Mackenzie