

Scientific Recommender: Análise Inicial Sobre Artigos Publicados em Conferências

Scientific Recommender: Initial Analysis On Articles Published in Conferences

Ingrid Pacheco¹, Eduardo Prata¹, Renan Parreira¹

¹Programa de Pós-Graduação em Informática – Instituto de Computação
Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brasil

{ingridpacheco}@dcc.ufrj.br, {edu.prata, renanparreira}@gmail.com

Abstract. *As a daily task, plenty of researchers submit articles to events and magazines hoping that their studies will be shared with more people. However, the frustration of being rejected can discourage them and overshadow their research. In the present work, analyses are performed on published articles over 10 conferences in order to recognize patterns through their histories and provide a foundation for the creation of a recommendation system that will assist them in the search of the perfect event.*

Keywords. *Recommendation Systems; Clustering; Published Articles; Data Science ; Data Analysis.*

Resumo. *Como uma tarefa rotineira, diversos pesquisadores submetem artigos a eventos e revistas com a esperança de terem seus estudos compartilhados com mais pessoas. Entretanto, a frustração ao ser rejeitado pode acabar o desestimulando e ofuscando uma brilhante pesquisa. Neste trabalho, são feitas análises em cima de artigos publicados em 10 conferências a fim de reconhecer padrões através de seus históricos e prover uma base para a criação de um sistema de recomendação que auxilie na busca pelo evento perfeito.*

Palavras-Chave. *Sistemas de Recomendação; Clusterização; Artigos Publicados; Ciência de Dados; Análise de Dados.*

1. Introdução

Um dos passos mais importantes na vida de um cientista, acadêmico, aluno ou pesquisador é a publicação de um artigo científico [Cuong et al. 2020]. O artigo dá mais importância e faz os trabalhos criados ganharem mais sentido pois o mesmo pode beneficiar outras pessoas agregando mais conhecimento em suas pesquisas.

Quando o pesquisador entende que seu trabalho já tem argumentos suficiente para responder as questões as quais ele se propõe a caracterizar e os problemas identificados possuem embasamento suficiente para serem resolvidos junto a suas hipóteses testadas, seria o momento perfeito para o mesmo partir para a publicação de seu artigo. Entretanto, um artigo científico deve ser relevante para os seus leitores e contribuir com novos conhecimentos, portanto, existe uma grande preocupação por parte destes de qual seria o evento ideal para a submissão, tentando diminuir a chance de negação.

Considerando que atualmente existem mais de 9585 conferências e 4152 *journals* de Ciência da Computação, é difícil para os autores saberem quais deveriam receber submissões, ainda mais considerando que submeter um artigo para a conferência errada na maior parte das vezes causa rejeição, atraso ou menos leituras no trabalho [Wang et al. 2018].

Algumas situações podem ser citadas como exemplos no que podem influenciar na rejeição de artigos e não se restringem a somente estes, como “o artigo não contribuir para novos conhecimentos na disciplina ou na aplicação de conhecimento” ou “o artigo não atender aos padrões éticos” ou até mesmo “o artigo estar mal escrito”. Ademais, um outro grande fator que pode contribuir para a rejeição de um artigo seria por parte da relevância oferecida para os seus leitores.

Para ajudar neste ponto, uma ferramenta de sistema de recomendação se faz fundamental para entender as necessidades dos pesquisadores e trazer para eles as melhores indicações de eventos nos quais os seus artigos teriam uma maior probabilidade de serem aceitos [Cuong et al. 2020]. Inclusive, este é um ponto tão relevante que diversos publicadores atualmente possuem suas próprias ferramentas de sistema de recomendação, como o *IEEE*¹, *Springer*² e *Elsevier*³.

Apesar do objetivo final ser criar um sistema de recomendação de conferências e *journals* baseado em ontologia, é preciso retroceder um pouco e entender como cada tópico se apresenta dentro de uma conferência, a fim de compreender os padrões delas, suas previsões e conseguir encontrar as melhores opções como resultado.

Portanto, neste contexto, este trabalho procura analisar os eventos e seus tópicos, formulando as seguintes perguntas de pesquisa a serem respondidas pelo mesmo:

- **P1** Quais são os 6 tópicos mais publicados nos eventos?
- **P2** Quais são os 6 tópicos mais publicados por ano (de 2018 a 2022) nestes eventos?
- **P3** Para cada tópico, quais as conferências que mais os publicam?
- **P4** Como foram as evoluções para cada tópico das conferências ao longo dos anos?
- **P5** Para cada tópico, em que categoria as conferências entram em relação a publicação dos mesmos?
- **P6** Quais as conferências que mais publicaram?
- **P7** Como foram as evoluções ao longo dos anos para as conferências?
- **P8** Para cada conferência, qual o tópico que mais aparece?

¹<https://publication-recommender.ieee.org/home>

²<https://journalsuggester.springer.com/>

³<https://journalfinder.elsevier.com/>

- **P9** Como foram as evoluções dos tópicos publicados por cada conferência ao longo dos anos?

Visando responder as questões de pesquisa de **P1** a **P9** elencadas e considerando a falta de um conjunto de dados único com as informações necessárias, o presente trabalho construiu um *dataset* com dados de artigos publicados em diversas conferências e realizou análises comparativas sobre eles para visualizar os atuais padrões. O conjunto de dados foi criado contendo 2.783 artigos da área de Ciência da Computação publicados em 10 conferências entre os anos de 2018 e 2022. Através dele foram realizadas análises de dados e criada uma clusterização baseada em seus tópicos principais.

Assim, foi possível encontrar os 6 tópicos mais publicados nessas conferências, tal como quais eram suas frequências em cada e sua evolução histórica. Ademais, A partir destes resultados, ficou perceptível quais os temas que mais estão em alta e com maior possibilidade de aceitação em uma submissão, assim como dependendo da temática, qual evento deveria ser priorizado.

O restante deste artigo está organizado da seguinte forma. Na seção 2, são discutidos os trabalhos relacionados a sistemas de recomendação para artigos e conferências. Na seção 3, é feita uma fundamentação teórica dos principais conceitos relacionados ao trabalho. Na seção 3.1 é apresentado o conceito principal de um sistema de recomendação. Na seção 3.2, por sua vez, é explicado o que é o método de clusterização, relevante para a análise posterior. Na seção 4, são apresentadas as informações sobre a criação dos conjuntos de dados. Na seção 5, são apresentadas as análises realizadas e os resultados obtidos. Por fim, na seção 6, são apresentadas as conclusões e os trabalhos futuros.

2. Trabalhos Relacionados

Nesta seção, são discutidos alguns trabalhos relacionados à sistemas de recomendação de conferências e *journals*, principalmente na área de Ciência da Computação.

Devido a necessidade de acelerar os processos de submissão de artigos, diversos sistemas de recomendação começaram a ser propostos, utilizando as mais diferentes metodologias e processos. Alguns destes são um novo modelo de recomendação de filtragem baseado em conteúdo (CBF) usando regressão qui-quadrado e *softmax* e outro de redes neurais.

2.1. Sistema baseado em Conteúdo

Em meados de 2020 um estudo emergiu relacionado a temática de sistemas de recomendação de conferências e *journals*, se apoiando em um *sistema online*⁴ que através do conteúdo inserido indicava um *top 3* ou *top 10* recomendados para submissão.

A proposta do grupo, apoiado por diversas instituições como *National Natural Science Foundation of China*⁵, National Basic Research Program of China e *Science Technology Development Project from Jilin Province*⁶, é de uma recomendação baseada em conteúdo, mais especificamente, o *abstract* ou o próprio manuscrito do artigo.

⁴<http://www.keaml.cn/prs/>

⁵https://www.nsfc.gov.cn/english/site_1/index.html

⁶<http://jlucqe.jlu.edu.cn/info/1201/6289.htm>

Ele se apresenta com um serviço web, e para garantir um bom e rápido funcionamento, possui um *web crawler* que constantemente atualiza o *set* de treinamento e o modelo de aprendizado.

Em relação a metodologia, é utilizado um modelo híbrido eficiente de *chi-square feature selection* e regressão *softmax*, a fim de garantir uma resposta online interativa. Entretanto, existe um limitante. O sistema só cobre 66 locais de publicação dentre mais de 5 bibliotecas digitais como Springer, IEEE, ACM, AAAI e SIAM.

Além disso, foram executados diversos testes com um *dataset* de 28 *journals* e 38 conferências, todas rankeadas A pela *China Computer Federation* (CCF). Como resultado eles obtiveram uma acurácia de 61.37% e o fato de que o sistema entregava uma resposta em 5 segundos na média [Wang et al. 2018], apesar de apresentar um resultado diferente ao ser testada por este trabalho.

2.2. Sistema baseado em redes neurais

Com o mesmo objetivo, mas diferente implementação, um estudo conduzido em esforço colaborativo entre *National Institute of Information and Communications Technology* e *University of Science, Vietnam National University at Ho Chi Minh City*, sugeriu o desenvolvimento de um sistema de recomendação utilizando técnicas de *natural language understanding*, combinando as *features* extraídas com outras *features* que sejam úteis [Cuong et al. 2020].

No trabalho mencionado, eles focam os esforços em dois domínios de pesquisa: Matemática e Ciência da Computação, e para realizar a entrega dos melhores resultados algumas etapas tem que ser realizadas. Para cada um dos tópicos, eles encontram a lista de *journals* e suas publicações, e usam elas para criar um *dataset* de treinamento e testagem da performance da ferramenta desenvolvida.

Um ponto relevante discutido acima é justamente a ampliação do campo de estudo, pois enquanto diversos trabalhos focam em somente uma área este tem a ambição de encontrar as melhores opções em duas, trazendo mais relevância para o estudo e podendo ser usufruído por mais pessoas.

Já em relação aos dados que são utilizados, eles são extraídos diretamente dos locais de publicação, e são coletadas todas as informações dos artigos como título, autores, resumo, palavras-chave, data de publicação, DOI, local de publicação, dentre outros. Além disso, para cada conferência e *journal* também é salvo o nome, escopo, índices, fator de impacto, o ranking, local e datas importantes. Ademais, eles mencionam o trabalho anterior feito pelo grupo no qual Wang faz parte [Wang et al. 2018]. Eles se beneficiam do dataset provido pelo mesmo que contém 14.012 pesquisas de Ciência da Computação divididas em 65 categorias diferentes. Já a parte dedicada a Matemática contém 392.894 artigos.

Para a parte de modelagem eles optaram por testar alguns modelos de recomendação e comparar suas performances para decidir o melhor a ser utilizado entre *handcrafted-features* e *deep-learning*.

Por fim, após a testagem eles integraram a ferramenta com o servidor, disponi-

bilizando uma plataforma na qual o usuário pode acessar, inserir os dados e receber a listagem dos locais de publicação recomendados. Entretanto, diferente do primeiro trabalho listado na atual seção, neste é necessário inserir alguns dados complementares além do resumo, como o título e a lista de possíveis palavras-chave. Ademais, eles também não disponibilizam o link da plataforma criada e não mencionam valores de performance antes citados como testados.

3. Fundamentação Teórica

Nesta seção, são apresentados alguns conceitos fundamentais que serão abordados ao longo do trabalho, incluindo o que é um sistema de recomendação e clusterização.

3.1. Sistema de Recomendação

Com o problema de sobrecarga de informação, os Sistemas de Recomendações se tornaram uma importante área de pesquisa ao longo dos anos, inicialmente chamados de Sistemas de Filtragem Colaborativa devido à utilização no primeiro sistema de recomendação Tapestry [GOLDBERG 1992], onde usuários indicavam documentos entre si, desenvolvido pela *Xerox*.

Os Sistemas de Recomendação são nada mais que classes de algoritmos com inteligência artificial que recomendam itens relevantes aos utilizadores, sendo utilizadas diferentes técnicas para a filtragem e ranqueamento da informação que observam vários fatores, como por exemplo, o histórico das interações dos utilizadores[]. Investigações associando os Sistema de Recomendação com novos recursos tecnológicos e o desenvolvimento de projetos e de produtos em diferentes tecnologias estão em crescimento nos últimos anos [Shao 2021]

A utilização mais comum nos dias de hoje está presente em sites de *e-commerce* ou canais de *streaming*, baseando suas recomendação pelos dados obtidos através do navegador, buscas e avaliações de produtos e serviços pela internet, de forma implícita.

Mas também podendo ser aplicada, de forma explícita, em outras áreas como a descrita neste artigo, onde será utilizada como ferramenta de recomendação para onde submeter um Artigo Científico buscando as melhores oportunidades, usando as palavras chaves do artigo como entrada para o Sistema de Recomendação será avaliado e analisado em um base de artigos.

3.2. Clusterização

Clusterização nada mais é do que um agrupamento de dados contáveis e quantificáveis buscando alguma similaridade. Sendo assim, é possível de ser utilizado no agrupamento de perfis ou de indivíduos com características em comum, visando representar padrões desses grupos de indivíduos, por exemplo, na identificação de um público alvo para uma campanha publicitária ou até mesmo para uma pesquisa de mercado, ou seja, a clusterização de clientes, clusterização de produtos, servidores, de dados.

Por sua vez, os *clusters* precisam de algoritmos com objetivo de expressarem tais critérios em forma de gráfico, como se os objetos a serem clusterizados se agrupassem

conforme uma maior ou menor distância em relação uns aos outros, de acordo com características em comum.

Por fim, os algoritmos mais usados são:

- **K-means Clusters:** Os *clusters* podem ser dispostos em gráficos hipotéticos, onde o centro de cada *cluster*, chamado de centroide, é calculado através da média dos valores. Desse modo o algoritmo buscará o centroide mais perto para cada integrante usando métricas de distância e atribuindo um ponto encontrado ao *cluster* mais próximo.
- **Clusterização hierárquica:** Na Clusterização hierárquica o algoritmo constrói uma hierarquia disposta em forma de dendrograma.

4. Construção do Conjunto de Dados

Nesta seção, será descrito o conjunto de dados gerado para realização das análises. Ele contém dados de diversos artigos publicados em 10 dos 11 principais eventos de Ciência da Computação em um intervalo de 5 anos (2018 até 2022). Posteriormente, o mesmo é utilizado para compreender a presença de cada temática nestes eventos e seu comparativo histórico ao longo dos anos.

Em relação a análise proposta, era necessária a construção de um conjunto de dados que contivesse todas as informações requeridas, como conferência, artigo, ano, tópicos, dentre outros. Para tal, foram extraídos 10 dos 11 principais eventos que tiveram publicações no tema Ciência da Computação. Estes vieram de uma *lista do departamento de Ciência da Computação da Universidade Cornell*⁷, e utilizaram como parâmetro a publicação de pelo menos 100 *papers* entre 2003 e 2013, com 5 citações cada durante os últimos 10 anos.

Os eventos utilizados para análise foram:

- *SOSP—ACM Symposium on Operating Systems Principles*
- *OSDI—Operating Systems Design and Implementation*
- *NDSS—Network and Distributed System Security Symposium*
- *MobiHoc—Mobile Ad Hoc Networking and Computing*
- *SIGCOMM—ACM SIGCOMM Conference*
- *SenSys—Conference On Embedded Networked Sensor Systems*
- *MOBICOM—Mobile Computing and Networking*
- *CIDR—Conference on Innovative Data Systems Research*
- *USENIX Security Symposium*
- *EUROCRYPT—Theory and Application of Cryptographic Techniques*

A partir das conferências selecionadas, o próximo passo foi coletar os artigos publicados em cada entre os anos de 2018 e 2022. Portanto, foi utilizada a fonte *dblp computer science bibliography*, que contém mais de 6 milhões de publicações e quase 6000 conferências mapeadas, com suas respectivas edições e conteúdos.

Entretanto, pelo fato de tal biblioteca digital não possuir uma *API* para facilitar a busca, foi necessário criar um *web scrapper* que passando por algumas propriedades da

⁷<https://www.cs.cornell.edu/andru/csconf.html>

página buscava cada uma das conferências mencionadas, depois entrava nas páginas de cada edição dentro do intervalo de tempo estabelecido e por fim, coletava as informações pertinentes aos seus artigos publicados. Vale ressaltar que a cada interação com a página alguns dados eram salvos, como o nome da conferência ao buscar sobre ela, o ano da edição na listagem de edições e dados sobre o artigo na listagem dos mesmos.

As rotas utilizadas e suas respectivas responsabilidades foram estabelecidas da seguinte forma:

- "<https://dblp.org/search?q=>": Consulta a biblioteca pelo nome da conferência para encontrar a listagem de edições dela.
- "<https://dblp.org/db/conf/confname>": Página com a listagem de edições da conferência. Ex: <https://dblp.org/db/conf/sosp/index.html>.
- "<https://dblp.org/db/conf/confname/confedition>": Página com a listagem de artigos publicados em determinada edição. Ex: <https://dblp.org/db/conf/sosp/sosp2021.html>.
- "<https://dl.acm.org/doi/articledoi>": Página com informações complementares sobre os artigos, como autores e *keywords*. Ex: <https://dl.acm.org/doi/10.1145/3477132.3483586>.

No fluxo em que o DOI não esteja atrelado ao artigo na listagem da edição, é feita uma busca auxiliar na *biblioteca do Semantic Scholar*⁸ utilizando o título do artigo, de forma a encontrar informações complementares sobre o mesmo.

Caso ainda assim não seja possível encontrar nem o *paperId* nem o DOI do artigo, os autores são coletados através de propriedades na própria listagem e o mesmo é salvo no *dataset* com as informações já existentes.

Se o DOI já tiver sido encontrado a partir do primeiro *scraper*, uma busca complementar é realizada na biblioteca do Semantic Scholar de forma a pegar os autores, tópicos e campo de estudo.

Por fim, já com o DOI encontrado na própria página ou pela consulta mencionada anteriormente, a página da *DL ACM*⁹ com informações complementares do artigo é consultada, com o intuito de obter tanto os autores (de mais fácil acesso) quanto mais *keywords* para fazer com que a análise seja mais completa possível. Todas as informações são posteriormente anexadas e salvas em um arquivo CSV, de forma a representar um artigo completo.

O conjunto de dados foi criado em 19 de Setembro de 2022 no formato CSV, contendo 2783 artigos publicadas em 10 eventos (SOSP, OSDI, NDSS, MobiHoc, SIGCOMM, SenSys, MOBICOM, CIDR, USENIX Security Symposium, EUROCRYPT) durante os anos de 2018 a 2022. Além disso, as informações foram armazenadas da forma apresentada na Tabela 1. Por fim, o conjunto de dados e o *script* em *Python* utilizado para o processo de geração do arquivo CSV foi disponibilizado para consulta e utilização em futuras pesquisas no *GitHub*¹⁰.

⁸<https://pypi.org/project/semanticscholar/>

⁹<https://dl.acm.org/doi/>

¹⁰<https://github.com/ingridpacheco/Scientific-Recommender>

Além disso, como o objetivo de melhorar o reuso do projeto, foi utilizada a biblioteca *prov*¹¹ para gerar a proveniência do código de geração de *dataset*. Todo o código e a imagem gerada pelo mesmo também estão disponíveis no repositório do Github.

Tabela 1. Informações os artigos no dataset

Atributo	Descrição
title	Título do artigo
paperId	PaperId do artigo disponibilizado no Semantic Scholar
doi	DOI do artigo, ou identificador único
authors	Autores que escreveram o artigo
publisher	Conferência na qual o artigo foi publicado
topics	Keywords dos artigos
fields_of_study	Campo geral de estudo do artigo
year	Ano que o artigo foi publicado

5. Análises e Resultados

Nesta seção, são detalhadas as análises e resultados encontrados ao longo do trabalho, tanto com relação as conferências quanto aos tópicos e seus comparativos históricos.

Baseado no conjunto de dados gerado, foi criado um novo *script*¹² para realizar as análises referentes as perguntas (especificamente de **P1** a **P9**) previamente estabelecidas (ver seção 1).

5.1. Caracterização do Dataset

Antes de efetivamente serem expostas as análises realizadas no *dataset*, algumas observações sobre ele merecem ser elucidadas, de forma que fique mais compreensível alguns resultados posteriores.

Apesar de o *web scraper* conseguir pegar uma boa parte dos dados dos artigos, nem todos eram disponibilizados nas páginas, o que acabou deixando certas brechas no *dataset* que tiveram que ser resolvidas manualmente nas análises. Outro ponto relevante é de que a análise inicial só foi feita em 10 conferências pois existiam limitações com relação a quantidade de requisições tanto na *API do Semantic Scholar* quanto ao site da DL ACM, que impossibilitavam uma raspagem maior de dados.

No total, além dos números apresentados anteriormente de 2783 artigos em 10 eventos entre 2018 e 2022, o *dataset* é composto por 3597 tópicos. Por fim, a quantidade de tópicos por conferência se faz da forma ilustrada na tabela 2.

¹¹<https://pypi.org/project/prov/>

¹²<https://github.com/ingridpacheco/Scientific-Recommender/blob/main/articleAnalysis.py>

Tabela 2. Quantidade de tópicos por conferência

Conferência	Quantidade de Tópicos
SOSP	383
OSDI	323
NDSS	1253
Mobihoc	628
SIGCOMM	485
SenSys	444
MOBICOM	702
CIDR	421
USENIX Security Symposium	1343
EUROCRYPT	395

5.2. Tópicos Que Mais Aparecem Em Artigos

Considerando o objetivo final de se criar um sistema de recomendação de eventos para pesquisadores de diversas áreas, é importante compreender quais os principais tópicos abordados em cada conferência para que assim as recomendações sejam mais certas usando as temáticas de cada pesquisador e o que os eventos mais prezam.

Dessa forma, e para responder as perguntas seguintes, foi-se criado um novo script¹³ focado na análise do *dataset*, utilizando como auxílio as bibliotecas: *numpy*¹⁴, *pandas*¹⁵, *matplotlib*¹⁶, *bokeh*¹⁷ e *plotly*¹⁸.

Para responder a **P1**, a biblioteca *pandas* foi utilizada na leitura do arquivo CSV criado com o *dataset*. Ela o transforma em um *DataFrame*¹⁹, estrutura de posterior fácil manipulação. Com este, um loop foi feito passando por cada linha (artigo) e um dicionário foi criado armazenando a quantidade de vezes que um tópico aparece (os tópicos tiveram que ser quebrados e armazenados um a um). Ao final da execução do mencionado trecho um dicionário da seguinte estrutura é disponibilizado:

```

1 {
2   'COMPUTER SYSTEMS ORGANIZATION' : 29,
3   'DEPENDABLE AND FAULT-TOLERANT SYSTEMS AND NETWORKS' : 18,
4   'SECURITY AND PRIVACY' : 25,
5   ...
6 }
```

A seguir, ele é ordenado de forma decrescente e os 6 tópicos com maior quantidade são recortados, gerando a listagem dos temas da tabela 3 e respondendo assim a **P1**:

¹³<https://github.com/ingridpacheco/Scientific-Recommender/blob/main/articleAnalysis.py>

¹⁴<https://numpy.org/>

¹⁵<https://pandas.pydata.org/>

¹⁶<https://matplotlib.org/>

¹⁷<http://bokeh.org/>

¹⁸<https://www.plotly.express/>

¹⁹<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>

Tabela 3. Quantidade de artigos de cada um dos 6 maiores tópicos

Tópico	Quantidade de artigos
'EXPERIMENT'	154
'ALGORITHM'	150
'COMPUTER SCIENCE'	131
'SENSOR'	114
'PROTOTYPE'	113
'MACHINE LEARNING'	107

Porém, o quadro acima considera o contexto completo de publicações, mas para compreender um pouco melhor as especificidades de cada ano, foi necessária a criação de um dicionário de anos, contendo a quantidade de publicações por tópico para aquele ano específico:

```

1 {
2   2022: {
3     'NETWORK CONGESTION': 1,
4     'ELASTICITY (DATA STORE)': 1,
5     'THROUGHPUT': 1,
6     'FREQUENCY RESPONSE': 1,
7     ...
8   }
9 }
```

No caso acima, no ano de 2022 teve uma publicação sobre '*NETWORK CONGESTION*', uma sobre '*ELASTICITY (DATA STORE)*', uma sobre '*THROUGHPUT*', uma sobre '*FREQUENCY RESPONSE*', dentre outras. Com o dicionário pronto, e para responder a **P2**, só foi preciso passar por cada ano ordenando de forma decrescente os tópicos para encontrar o *top 6* de cada um, referido na tabela 4:

Tabela 4. Top 6 Tópicos Publicados Em Cada Ano

Ano	Top 6 Tópicos
'2018'	'EXPERIMENT', 'SENSOR', 'PROTOTYPE', 'ALGORITHM', 'OVERHEAD (COMPUTING)', 'THROUGHPUT'
'2019'	'ALGORITHM', 'EXPERIMENT', 'PROTOTYPE', 'MACHINE LEARNING', 'SENSOR', 'SCALABILITY'
'2020'	'COMPUTER SCIENCE', 'ALGORITHM', 'EXPERIMENT', 'NETWORKS', 'MACHINE LEARNING', 'ADVERSARY (CRYPTOGRAPHY)'
'2021'	'COMPUTER SCIENCE', 'NETWORKS', 'SOFTWARE AND ITS ENGINEERING', 'COMPUTER SYSTEMS ORGANIZATION', 'SOFTWARE ORGANIZATION AND PROPERTIES', 'INFORMATION SYSTEMS'
'2022'	'COMPUTER SCIENCE', 'NETWORK CONGESTION', 'ELASTICITY (DATA STORE)', 'THROUGHPUT', 'FREQUENCY RESPONSE', 'FLOW'

5.3. Eventos Que Mais Publicam Cada Tópico

A fim de compreender para cada um destes 6 tópicos sua distribuição nos 10 eventos, outros dicionários tiveram de ser criados. O primeiro deles armazenava a correlação da

quantidade de artigos publicados de cada tópico em cada evento em cada ano, como mostra a estrutura seguinte:

```

1 {
2   'NP-HARDNESS' : {
3     'MobiHoc' : {
4       2020: 2,
5       2019: 2,
6       2018: 2
7     },
8     'MOBICOM' : {
9       2018: 1
10    }
11  },
12  'ONLINE LEARNING SETTINGS' : {
13    'MobiHoc' : {
14      2021: 1
15    }
16  }
17  ...
18 }

```

No caso acima, o tópico '*NP-HARDNESS*' teve na conferência MobiHoc: 2 publicações em 2020, 2 no ano de 2019 e 2 em 2018. Já na MOBICOM, teve 1 publicação em 2018.

O segundo *dataset*, por sua vez, armazenava a correlação oposta, ou seja, a quantidade de artigos publicados de um determinado tópico em cada ano em cada evento, da forma:

```

1 {
2   'TROJAN HORSE (COMPUTING)' : {
3     2018: {
4       'NDSS' : 1,
5       'USENIX Security Symposium' : 1
6     },
7     2021: {
8       'USENIX Security Symposium' : 1
9     },
10    2020: {
11      'USENIX Security Symposium' : 1
12    },
13    2019: {
14      'USENIX Security Symposium' : 3
15    }
16  },
17  ...

```

Ou seja, o tópico '*TROJAN HORSE (COMPUTING)*' em 2018 teve 1 publicação no NDSS e 1 no USENIX Security Symposium, enquanto em 2021 só teve 1 publicação no último.

Com os dois dicionários já criados, para cada tópico do *top 6* o processo foi pegar a quantidade por ano de cada evento, transformar em um *Dataframe*, agregar a soma de cada evento numa outra variável, fazer um recorte das quantidades do último ano e enfim, criar um *Dataframe* com as colunas: '*Publisher*' (evento), '*Total*' e '*LastYear Qty*' (Quantidade de publicações de determinado tópico no evento específico no último ano disponível).

A partir da estrutura criada acima foi possível tanto compreender como cada tema era abordado de uma maneira geral por cada conferência, respondendo a **P3** quanto fazer uma comparação histórica ao longo dos anos, respondendo a **P4**, para entender em quais ele mais aumentou, se manteve o mesmo ou diminuiu, como ilustrado na figura 5.3 para o caso do tópico '*EXPERIMENT*'.

Todos os gráficos produzidos para os outros 5 tópicos estarão no apêndice 1 do presente trabalho para futuras consultas. Porém, uma resumida análise sobre cada um deles se encontra na tabela 6:

Tabela 5. Top 3 Conferências Que Mais Publicam Cada Tópico Com Suas Respectivas Quantidades

Tópico	Primeiro lugar	Segundo lugar	Terceiro lugar
EXPERIMENT	USENIX Security Symposium (38)	NDSS (31)	MOBICOM (31)
ALGORITHM	USENIX Security Symposium (33)	MOBICOM (26)	NDSS (24)
COMPUTER SCIENCE	USENIX Security Symposium (56)	NDSS (25)	CIDR (24)
SENSOR	USENIX Security Symposium (33)	SenSys (28)	MOBICOM (24)
PROTOTYPE	NDSS (29)	MOBICOM (28)	USENIX Security Symposium (16)
MACHINE LEARNING	USENIX Security Symposium (29)	NDSS (19)	MOBICOM (15)

Ademais, cabe levantar as seguintes observações:

- Apesar de não estar entre as conferências que mais publicam sobre EXPERIMENT, o SIGCOMM foi o único que teve alguma publicação em 2022 do tópico. Por fim, o único evento que apresentou um aumento constante entre 2018, 2019 e 2020 foi o NDSS, mesmo a quantidade tendo diminuído em 2021, sendo o único a ter publicado artigo sobre o tema neste ano.
- A única conferência que publicou artigo sobre ALGORITHM em 2022 foi EUROCRYPT.

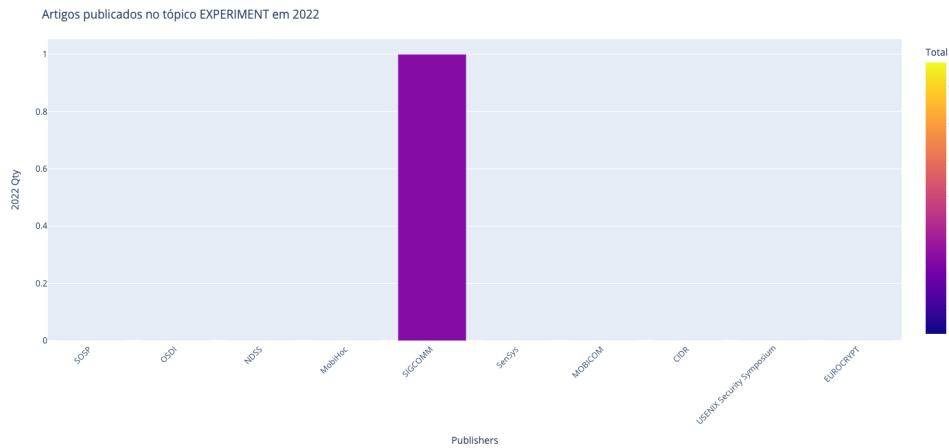


Figura 1. Artigos publicados sobre EXPERIMENT em 2022

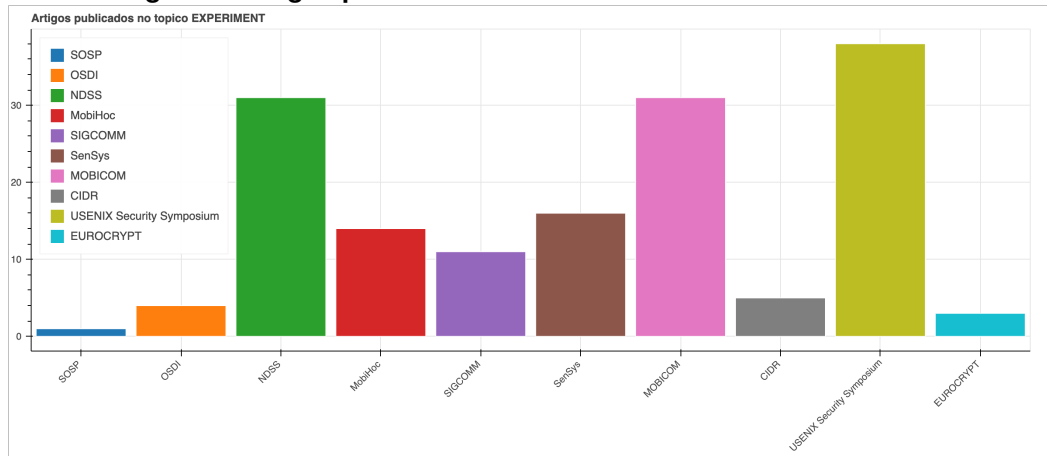


Figura 2. Artigos publicados sobre EXPERIMENT

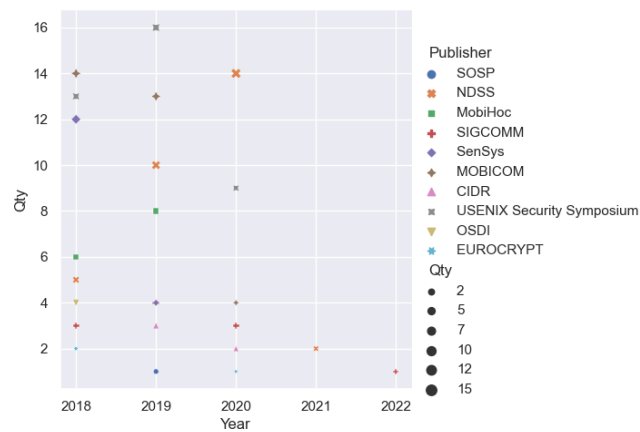


Figura 3. Comparativo hist rico dos eventos para artigos sobre EXPERIMENT

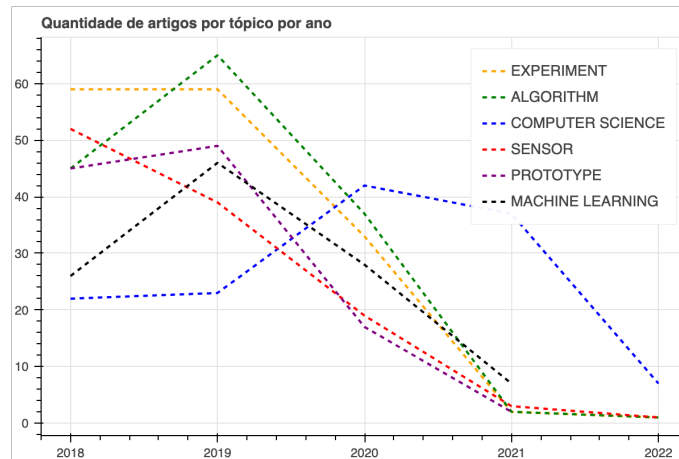


Figura 4. Artigos publicados em cada tópico por ano

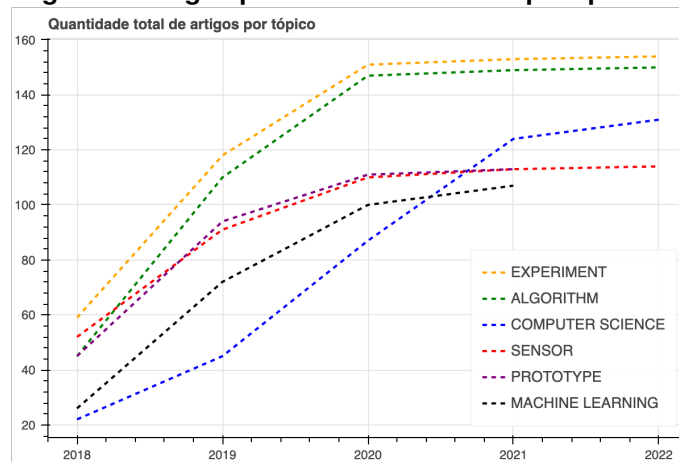


Figura 5. Artigos publicados em cada tópico

Além disso, dá para perceber através das figuras 4 e 5 que alguns temas ainda não tiveram publicações em 2022, como MACHINE LEARNING e PROTOTYPE. Já com relação a quantidade total ao longo dos anos, o tema COMPUTER SCIENCE teve um bom crescimento entre 2019 e 2021, enquanto outros se mantiveram mais estáveis.

Apesar de ser importante compreender de maneira mais específica as publicações das conferências, também é interessante ter um olhar mais genérico para as distribuições das publicações deles sobre um tópico em relação as publicações totais.

Para isso, foram criados dois dicionários, um de publicações de conferências por tópico, e outro de publicações de conferências por ano, representados abaixo, respectivamente:

```

1 {
2   'SOSP' : {
3     'COMPUTER SYSTEMS ORGANIZATION' : 23,
4     'DEPENDABLE AND FAULT-TOLERANT SYSTEMS AND NETWORKS' : 18,
```

```

5      'SECURITY AND PRIVACY' : 12,
6      'DATABASE AND STORAGE SECURITY' : 2
7      ...
8  }
9  }

```

```

1  {
2      'SOSP' : {
3          2021: 54,
4          2019: 38
5      },
6      'OSDI' : {
7          2021: 31,
8          2020: 70,
9          2018: 47
10     }
11 }

```

No primeiro dicionário é possível de se ver que o SOSP teve 23 publicações de 'COMPUTER SYSTEMS ORGANIZATION' e 12 de 'SECURITY AND PRIVACY'. Em seguida, percebe-se que o mesmo teve 54 publicações em 2021 e 38 em 2019.

Já com os dicionários disponíveis, para cada um dos 6 tópicos, pegou-se a quantidade total de publicações por conferência e foi comparado com a quantidade de publicações do tópico na conferência, usando o algoritmo de clusterização *K-Means* por tópico.

Rodando o algoritmo para encontrar a quantidade ideal de *clusters*, foi-se gerada a figura 6, que resultou em 3 como a quantidade adequada.

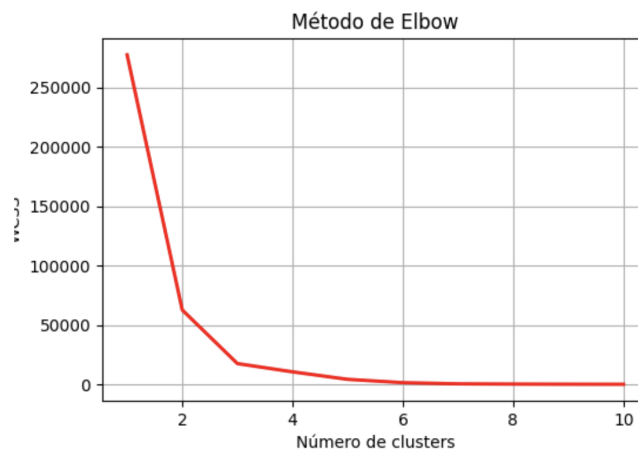


Figura 6. Quantidade de cluster ideal para clusterização

Com os dados definidos e a quantidade de *clusters* encontrada, foi possível gerar a clusterização para cada tópico e responder a **P5**, assim como representado pela do

EXPERIMENT ilustrado na figura 7. As imagens respectivas aos outros tópicos estão no apêndice 1 do trabalho.

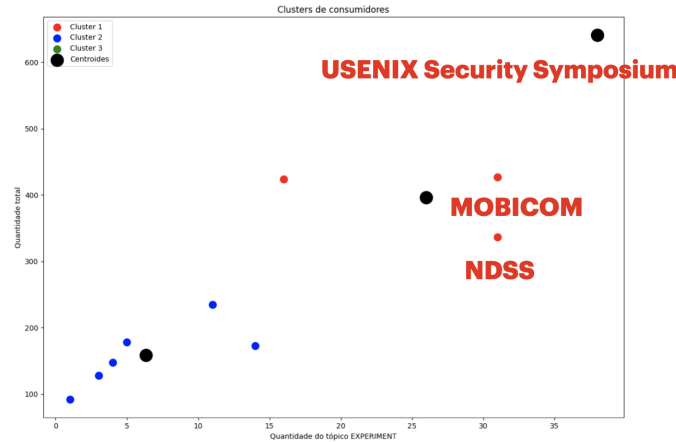


Figura 7. Clusterização do tópico EXPERIMENT

Resumidamente, as descobertas do processo de clusterização foram que apesar do USENIX Security Symposium ter a maior quantidade total de artigos publicados, em alguns casos outras opções são melhores, como na lista abaixo:

- Para o EXPERIMENT, MOBICOM e NDSS apresentam uma correlação bem mais interessante de publicações totais e do tópico mencionado, pois apesar de terem menos no total, a quantidade do tópico chega bem perto do USENIX Security Symposium, o que as torna ótimas conferências a serem recomendadas para o definido tema.
- Para COMPUTER SCIENCE, NDSS, CIDR e OSDI são boas opções também, apesar do USENIX Security Symposium, nesse caso, ter uma quantidade maior do tópico.
- Para ALGORITHM, MOBICOM, NDSS e SIGCOM são boas opções também e chegam bem perto de USENIX Security Symposium.
- Para SENSOR, MOBICOM e SenSys são ótimas opções com uma relação muito boa entre publicações totais e do tópico.
- Para PROTOTYPE, MOBICOM e NDSS são opções muito melhores que USENIX Security Symposium, tanto considerando que a quantidade de publicações do tópico é maior, quanto a de publicações totais menor, o que faz com que a relação seja mais interessante nesse caso.
- Para MACHINE LEARNING, MOBICOM e NDSS são boas opções, mas USENIX Security Symposium ainda possui mais publicações do tópico.

5.4. Eventos Que Mais Publicam Artigos

Além de analisar os tópicos com maior profundidade, também é importante entender o comportamento das conferências com relação as suas publicações, para assim recomendar o melhor deles baseado no seu histórico e padrões de publicação.

Para calcular as quantidades de publicações de cada conferência, e responder a **P6**, foi utilizado o dicionário anteriormente mencionado de publicação de conferências por ano e somadas as publicações de todos os anos, de forma a resultar na quantidade total por conferência, como ilustrado na figura 8:

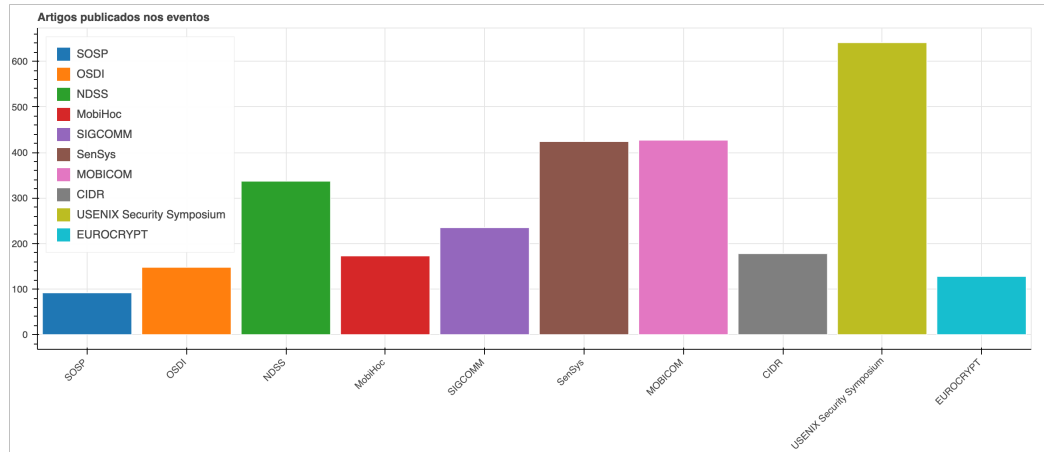


Figura 8. Artigos publicados por conferência

De acordo com a imagem acima, percebe-se que USENIX Security Symposium foi a conferência com mais publicações, seguida pelo MOBICOM e SenSys, com 641, 427 e 424 publicações respectivamente.

Apesar disso, também é relevante entender o padrão ao longo dos anos de publicações, se aumentou, diminuiu, ou se manteve o mesmo. Para isso, e para responder a **P7**, foi utilizado novamente o dicionário de publicações das conferências por ano, mas ao invés de simplesmente somar tudo de uma vez, foi-se sendo guardado o registro total ao longo dos anos, de forma a conseguir gerar as figuras 9 e 10:

O padrão seguiu o comentado na figura anterior, mantendo USENIX Security Symposium, MOBICOM e SenSys no topo, mas é importante ver como alguns deles aumentaram a quantidade de publicações nos últimos anos enquanto outros decaíram, como OSDI e SenSys.

5.5. Tópicos Que São Mais Publicados Por Cada Conferência

Para descobrir quais os tópicos mais publicados por cada conferência, e assim responder a **P8**, alguns processamentos tiveram de ser feitos, como pegar o dicionário de publicações de conferências por tópico e ordenar para pegar os 6 com os maiores valores. Depois disso, se usou a quantidade total de publicações de um tópico com a quantidade dele em cada conferência, para assim, possibilitar uma comparação de valores. Dessa forma, foi possível encontrar o *top 6* de tópicos para cada conferência e compreender a sua relevância na quantidade total dos tópicos.

Foi, então, gerada uma imagem para cada conferência com os seus tópicos mais publicados, como está ilustrado no caso do evento SOSP na figura 11. Para as outras conferências, suas figuras estão no apêndice 1 do presente trabalho.

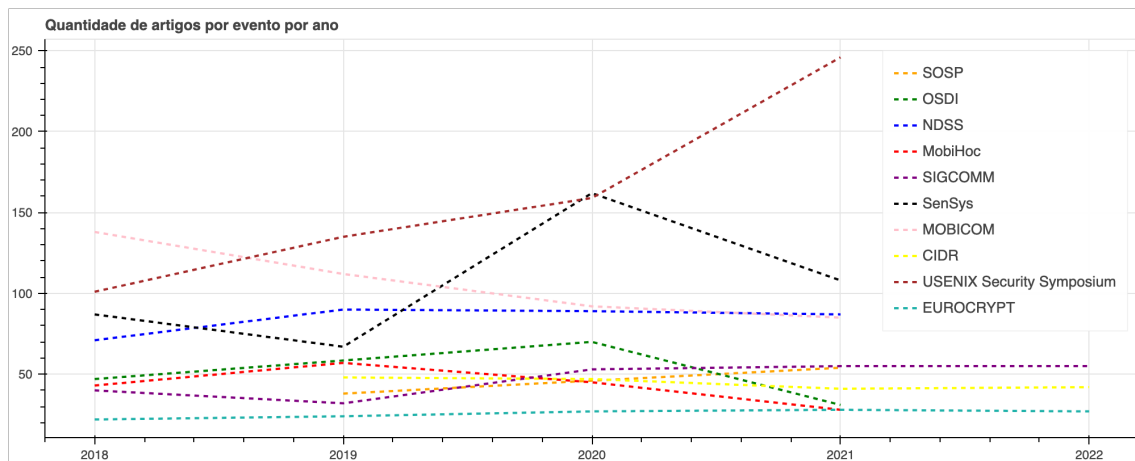


Figura 9. Artigos publicados nas conferências por ano

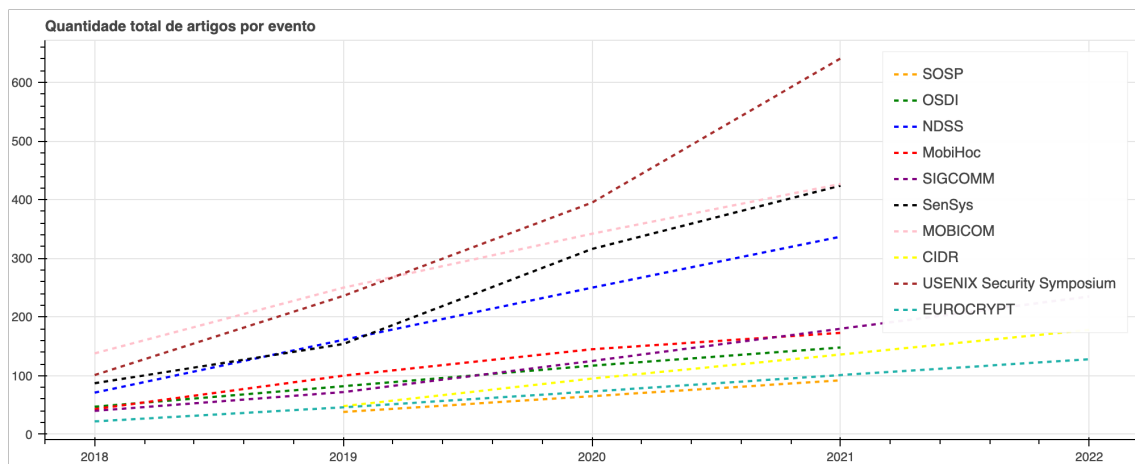


Figura 10. Artigos publicados nas conferências no total

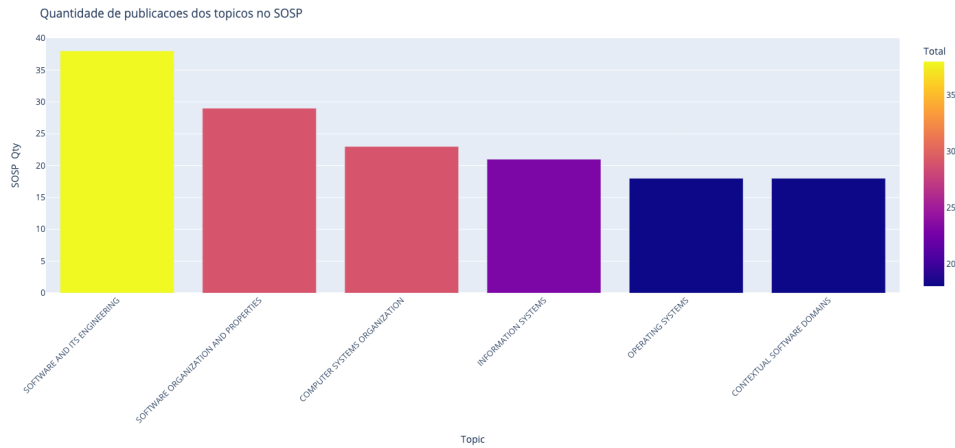


Figura 11. Tópicos mais publicados no SOSP

Os 6 tópicos mais publicados por cada conferência são:

Tabela 6. Top 6 Tópicos Mais Publicados Por Cada Conferência

Conferência	Top 6 Tópicos E Suas Quantidades
SOSP	SOFTWARE AND ITS ENGINEERING (38), SOFTWARE ORGANIZATION AND PROPERTIES (29), COMPUTER SYSTEMS ORGANIZATION (23), INFORMATION SYSTEMS (21), OPERATING SYSTEMS (18) e CONTEXTUAL SOFTWARE DOMAINS (18)
OSDI	COMPUTER SCIENCE (19), SERVER (COMPUTING) (9), THROUGHPUT (9), CENTRAL PROCESSING UNIT (8), DISTRIBUTED COMPUTING (7) e COMPUTATION (7)
NDSS	OVERHEAD (COMPUTING) (43), ADVERSARY (CRYPTOGRAPHY) (32), EXPERIMENT (31), MALWARE (30), PROTOTYPE (29) e COMPUTER SCIENCE (25)
MobiHoc	NETWORKS (64), NETWORK PERFORMANCE EVALUATION (31), NETWORK TYPES (28), SIMULATION (25), NETWORK PERFORMANCE ANALYSIS (21) e NETWORK ALGORITHMS (21)
SIGCOMM	ALGORITHM (21), NETWORK PACKET (18), THROUGHPUT (13), NETWORK CONGESTION (12), EXPERIMENT (11) e SOFTWARE DEPLOYMENT (10)
SenSys	SENSOR (28), EXPERIMENT (16), SOFTWARE DEPLOYMENT (15), PROTOTYPE (14), REAL-TIME CLOCK (13) e SMARTPHONE (13)
MOBICOM	EXPERIMENT (31), PROTOTYPE (28), ALGORITHM (26), SENSOR (24), SMARTPHONE (22) e THROUGHPUT (20)
CIDR	COMPUTER SCIENCE (24), DATABASE (22), MATHEMATICAL OPTIMIZATION (15), MACHINE LEARNING (12), QUERY OPTIMIZATION (9) e PROTOTYPE (8)
USENIX Security Symposium	COMPUTER SCIENCE (56), EXPERIMENT (38), ALGORITHM (33), SENSOR (33), OVERHEAD (COMPUTING) (31) e VULNERABILITY (COMPUTING) (31)
EUROCRYPT	CRYPTOGRAPHY (22), TIME COMPLEXITY (13), COMPUTATION (13), ALGORITHM (11), IBM NOTES (9) e POLYNOMIAL (9)

Entretanto, devido a relação entre quantidade total e na devida conferência, algumas observações podem ser feitas:

- No SOSP, os tópicos OPERATING SYSTEMS (18) e CONTEXTUAL SOFTWARE DOMAINS (18) tem a mesma quantidade total e na conferência, o que significa que essa conferência é forte nesse tema, portanto ela possivelmente deveria estar na frente na recomendação por esses tópicos.
- No OSDI, a relação no tópico DISTRIBUTED COMPUTING é a melhor, considerando que este tem um total de 22 artigos e 7 deles estão nessa conferência.
- Já para o NDSS, o tema OVERHEAD (COMPUTING) possui uma boa correlação de quantidade total e no evento, assim como o tema MALWARE.
- Para o MobiHoc, apesar do tópico NETWORKS estar bem, outros dois interessantes são NETWORK PERFORMANCE ANALYSIS (21) e NETWORK ALGORITHMS (21), que possuem a mesma quantidade total e na conferência.
- Para SIGCOMM, NETWORK CONGESTION é o que possui a melhor correlação.
- Para o SenSys, o tópico REAL-TIME CLOCK possui 40 artigos no total e 13 na conferência, sendo uma das melhores relações para o evento.
- Para MOBICOM, tanto SMARTPHONE (22) quanto THROUGHPUT (20) são interessantes.
- Para CIDR, DATABASE e QUERY OPTIMIZATION possuem ótimas relações com 31 e 10 no total e 22 e 9 na conferência, respectivamente.
- Para USENIX Security Symposium, os tópicos COMPUTER SCIENCE e VULNERABILITY (COMPUTING) trazem ótimos números.
- Para EUROCRYPT, os tópicos CRYPTOGRAPHY e POLYNOMIAL tem as melhores relações e se tornam tópicos fortes para recomendação da conferência.

Portanto, é muito importante olhar para todo o contexto e não somente para números absolutos, a fim de conseguir encontrar melhores correlações para uma possível recomendação.

Por fim, para responder a **P9** e entender a evolução dos tópicos nas conferências ao longo dos anos, foi necessária a criação de um dicionário de publicações das conferência por ano e tópico, como na demonstração a seguir:

```

1 {
2   'MOBICOM' : {
3     2020: {
4       'LOCATION (GEOGRAPHY)' : 1,
5       'DUPLEX (TELECOMMUNICATIONS)' : 1,
6       'DEFINITION' : 1,
7       'PHASED ARRAY' : 2,
8       'EXPERIMENT' : 4,
9       'FEEDBACK' : 1,
10      'NETWORK PERFORMANCE' : 1,
11      'APPROXIMATION ALGORITHM' : 1,
12      'PROGRAMMING PARADIGM' : 1,
13      'DOWNTIME' : 2,
14      'IBM NOTES' : 1,
15      'BACKSCATTER (EMAIL)' : 1,

```

16 . . .
 17 }
 18 }
 19 }

Com o dicionário, só foi necessário pegar os 6 tópicos com mais publicações para cada conferência e suas quantidades por ano, resultando num histograma de suas publicações ao longo do tempo. A figura 12 ilustra o histograma da conferência SOSP com relação aos seus 6 tópicos com mais publicações, já mencionado antes. Os histogramas das demais conferências estão no apêndice 1 do trabalho.

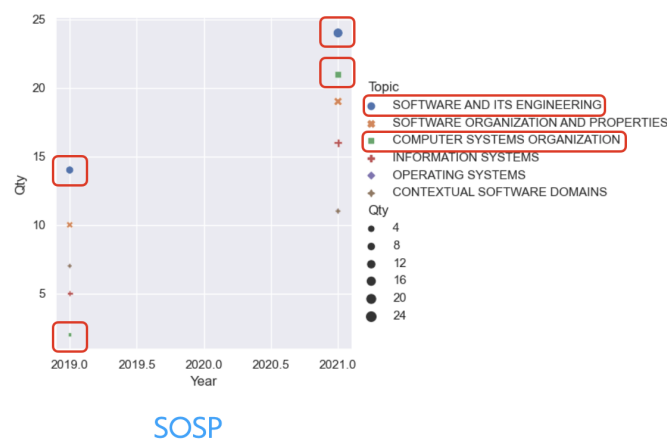


Figura 12. Histograma de tópicos do SOSP

De forma resumida, as descobertas dos histogramas foram as seguintes:

- Para SOSP, os tópicos SOFTWARE AND ITS ENGINEERING e COMPUTER SYSTEMS ORGANIZATION apareceram nos anos de 2019 e 2021 (únicos com valores nesse evento) e tiveram uma evolução considerável, se tornando os dois com mais publicações em 2021.
- Para OSDI, COMPUTER SCIENCE foi o único tópico que apareceu nos três anos com valores, apesar de ter diminuído em 2021.
- Para o NDSS, os dois tópicos com mais publicações em 2021 foram EXPERIMENT e COMPUTER SCIENCE, e como os dois aparecem em todas as edições da conferência, se tornam fortes tópicos para ela.
- Para o MobiHoc, o tópico NETWORKS é o que tem as maiores publicações nos dois últimos anos e o NETWORK PERFORMANCE ANALYSIS é o que aparece na maior quantidade de edições, 3 de 4.
- Para SIGCOMM, THROUGHPUT é o que aparece com a maior quantidade em 2021 e se repete em duas edições, e EXPERIMENT se repete em três edições, sendo o único tópico a ter publicações em 2022.
- Para o SenSys, o tópico SENSOR é o que possui os maiores valores para 2018 e 2019, enquanto SOFTWARE DEPLOYMENT aparece em segundo em 2019.

- Para MOBICOM, só três temas aparecem nas três edições de 2018, 2019 e 2020, EXPERIMENT, ALGORITHM e SMARTPHONE, tendo os dois primeiros maiores quantidades de publicações na conferência.
- Para CIDR, o tópico que aparece na maior parte das edições, em 2019, 2021 e 2022 é COMPUTER SCIENCE.
- Para USENIX Security Symposium, os dois tópicos que mais aparecem são COMPUTER SCIENCE, em 2021, 2020 e 2018, e OVERHEAD (COMPUTING), em 2021, 2020 e 2019. Cabe ressaltar que em 2021 o tópico COMPUTER SCIENCE teve a maior quantidade de publicações.
- Para EUROCRYPT, o tópico CRYPTOGRAPHY foi o único que apareceu em todas as edições (2018 a 2021), mas em 2021 o tópico ALGORITHM foi quem teve a maior quantidade de publicações.

Portanto, através das análises foi possível compreender de forma mais completa os padrões e singularidades de cada conferência e tópico, sabendo onde mais aparecem e seus históricos ao longo dos anos. Isso é apenas o começo de um longa possível descoberta sobre os artigos publicados.

Por fim, cabe ressaltar que com o objetivo de melhorar o reuso do trabalho também foi inserida proveniência no código, e tanto ele quanto a *imagem gerada*²⁰ estão disponíveis no Github.

6. Conclusões

Como uma das atividades mais comuns na vida dos pesquisadores, submeter artigos a eventos e revistas pode ser cansativo se o mesmo não souber as melhores plataformas e as que vão valorizar as suas principais descobertas. Além de longas descrições e formulários complexos, muito se gasta na expectativa de ser aceito e poder ter uma publicação compartilhada com diversas outras pessoas. Infelizmente, devido a grande diversidade de conferências e eventos muito potencial é descartado no meio do caminho, e por isso, uma ferramenta que fosse capaz de recomendar às pessoas o local com maior chance de aceitação diminuiria a fricção e busca, a fim de ajudar mais cientistas a terem seus estudos abertos com toda a comunidade.

Mediante o exposto, e considerando este trabalho como uma etapa inicial essencial para o objetivo final, é possível afirmar que ele, através das análises propostas, trouxe luz a forma como algumas conferências estão se portando a certa de temas centras, como protótipos, algoritmos e *machine learning*. Ademais, foi possível identificar padrões históricos que podem servir para futuramente uma previsão de quantidade de artigos publicados de um tema para cada evento, considerando como eles se portaram no passado. Além disso, este artigo pode ser útil para fomentar a discussão sobre quais temas deveriam ser mais publicados ou quais não estão recebendo atenção suficiente em prol de outros, elucidando buracos no conhecimento que por vezes podem estar obscuros.

No entanto, durante o desenvolvimento deste trabalho, alguns desafios foram encontrados. O primeiro foi a falta de um conjunto de dados único com todas as informa-

²⁰<https://github.com/ingridpacheco/Scientific-Recommender/blob/main/Proveniencia/articleAnalysis-prov.png>

ções, sendo necessária a criação de um *script* em *Python* para, coletando os dados de diferentes fontes, fazê-lo. Além, este trabalho teve algumas limitações, como a quantidade de 100 requisições para a API do Semantic Scholar a cada 5 minutos e o fato do site do DL ACM bloquear o IP do computador após alguns *scrapings* por achar que seria um ataque *hacker*. Devido a estes fatos a pesquisa se limitou a 10 conferências, e para tentar driblar essas questões teve que usada uma função de delay no código e uma VPN externa, respectivamente. Outro ponto que cabe ressaltar é que cada evento na página do *dblp* tinha uma estrutura distinta, então alguns artigos ficaram com dados faltantes, pois o *scraping* procurava por certos padrões que não existiam.

Como trabalhos futuros, pretende-se ampliar essas análises a uma maior quantidade de conferências e a um intervalo maior de tempo para ter um contexto histórico mais completo. Outro ponto interessante é o de expandir as análises para questões relacionadas ao países dos artigos que os eventos mais publicam e se existe algum relação entre os artigos publicados por uma mesma conferência. As análises propostas neste trabalho servem como munição inicial para diversas outras abordagens, além do próprio sistema de recomendação do qual ele é base. Nesse site, pode-se utilizar também o campo de estudo do autor para já fazer uma busca por conferências direcionada ao invés de uma busca genérica. Além disso, também seria interessante receber do autor um evento que ele tenha como favorito para tentar encontrar conferências que se assimilem a ela, em que-sito de tópicos publicados ou outros padrões. Finalmente, é muito importante encontrar uma forma de trazer todos os dados dos artigos automaticamente, sendo pelo uso de outra *API* ou algum site que contenha as informações completas, que atualmente se encontram indisponíveis para alguns casos.

Referências

- Cuong, D. V., Nguyen, D. H., Huynh, S., Huynh, P., Gurrin, C., Dao, M.-S., Dang-Nguyen, D.-T., e Nguyen, B. T. (2020). A framework for paper submission recommendation system. In *Proceedings of the 2020 International Conference on Multimedia Retrieval, ICMR '20*, page 393–396, New York, NY, USA. Association for Computing Machinery.
- GOLDBERG, D. e. a. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35:61–70.
- Shao, B., L. X. . B. G. (2021). A survey of research hotspots and frontier trends of recommendation systems from the perspective of knowledge graph. *Expert Systems with Applications*, 165.
- Wang, D., Liang, Y., Xu, D., Feng, X., e Guan, R. (2018). A content-based recommender system for computer science publications. *Knowledge-Based Systems*, 157:1–9.

1. Apêndices

Nesta seção serão colocadas as imagens complementares indicadas no texto principal

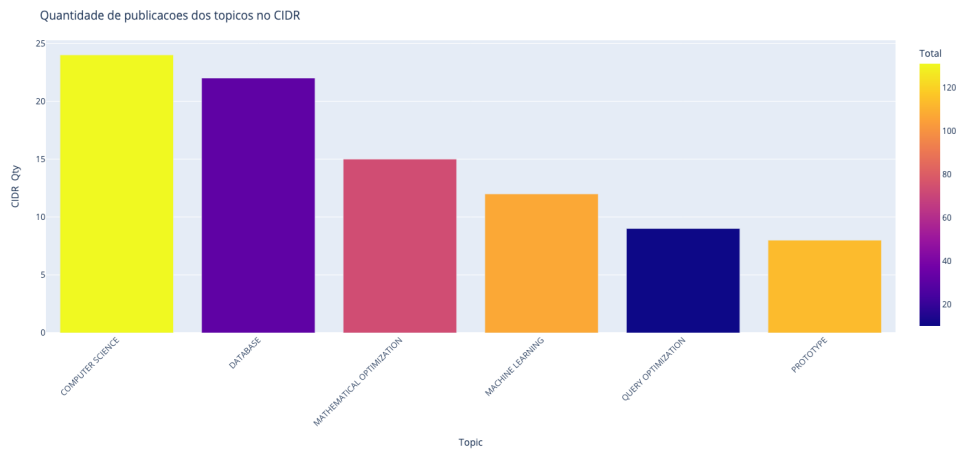


Figura 13. Tópicos mais publicados no CIDR

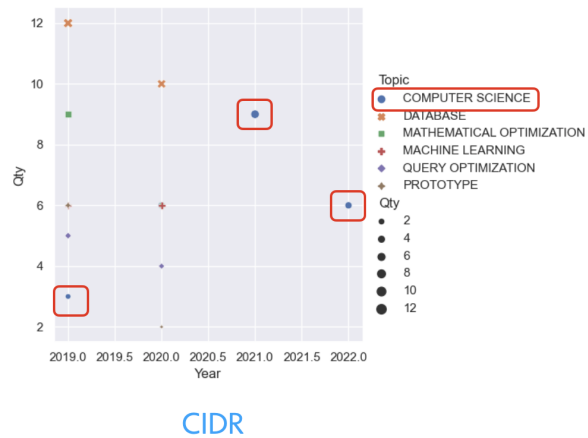


Figura 14. Histograma de tópicos no CIDR

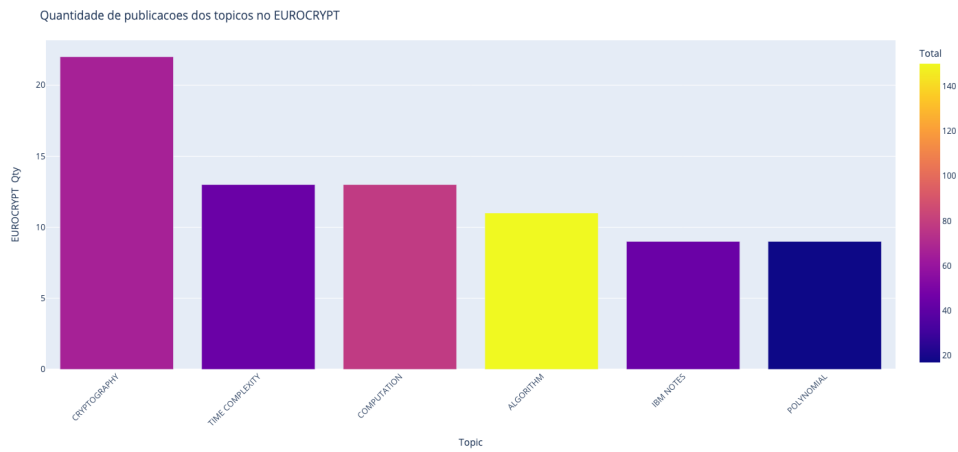


Figura 15. Tópicos mais publicados no EUROCRYPT



EUROCRYPT

Figura 16. Histograma de tópicos no EUROCRYPT

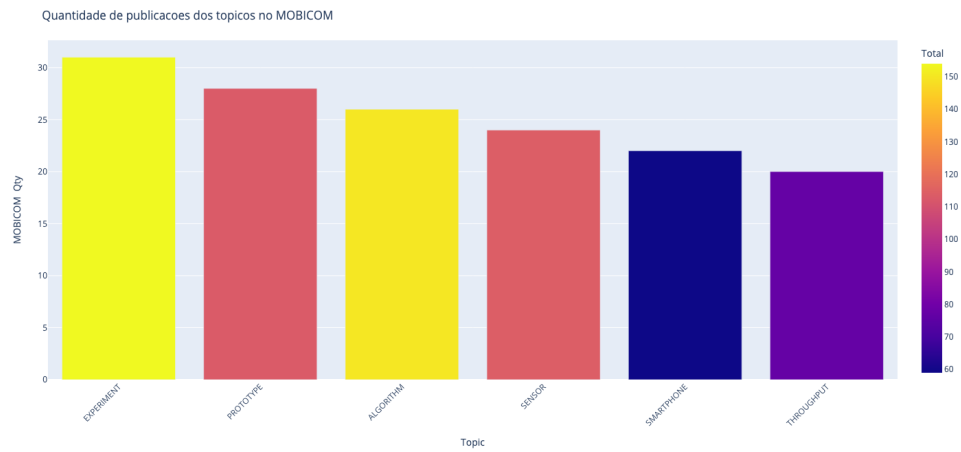
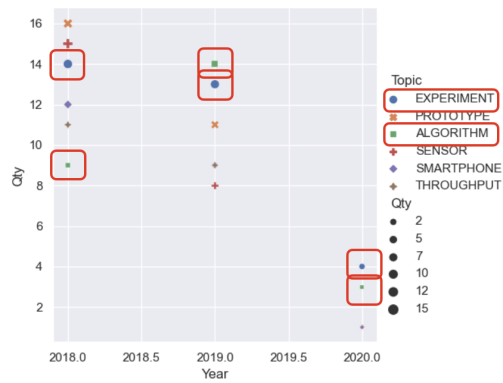


Figura 17. Tópicos mais publicados no MOBICOM



MOBICOM

Figura 18. Histograma de tópicos no MOBICOM

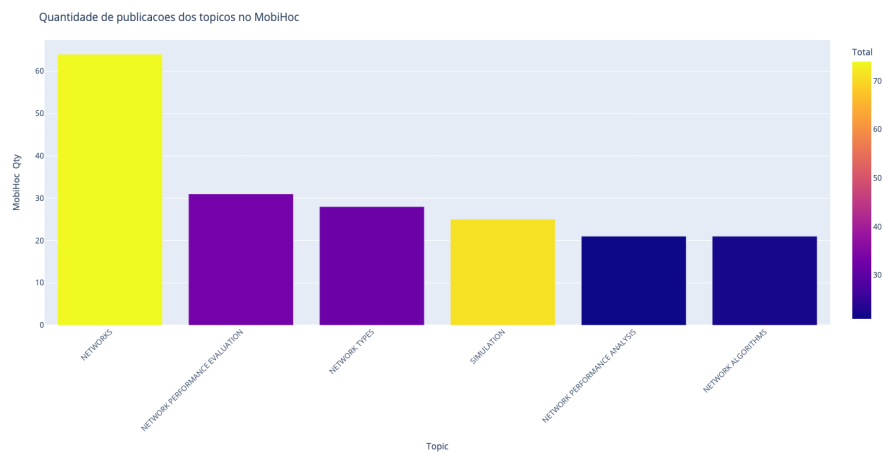


Figura 19. Tópicos mais publicados no MobiHoc



MobiHoc

Figura 20. Histograma de tópicos no MobiHoc

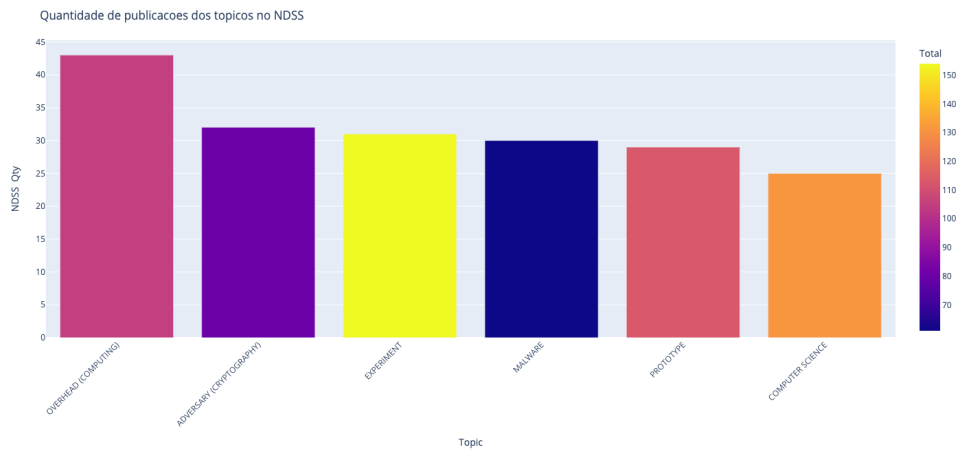


Figura 21. Tópicos mais publicados no NDSS



NDSS

Figura 22. Histograma de tópicos no NDSS

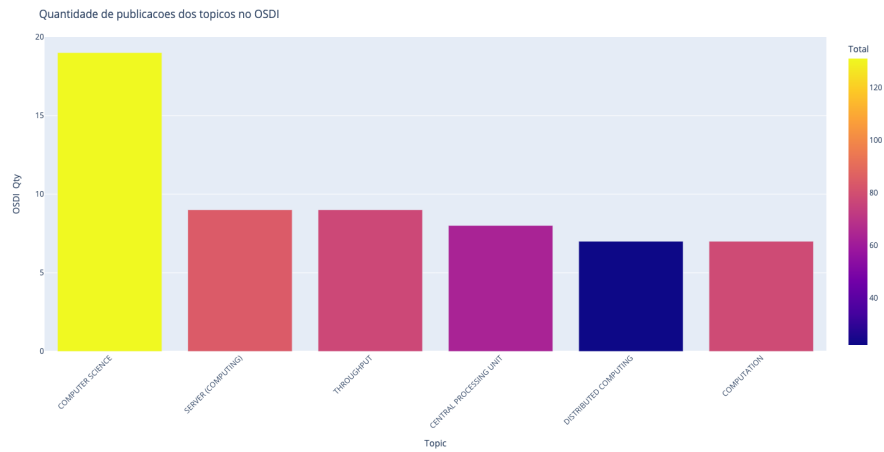


Figura 23. Tópicos mais publicados no OSDI

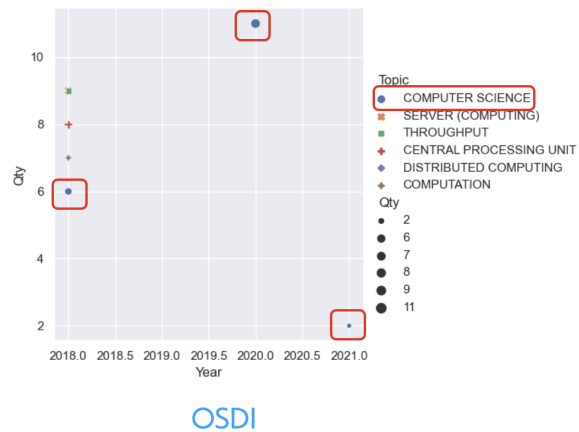


Figura 24. Histograma de tópicos no OSDI

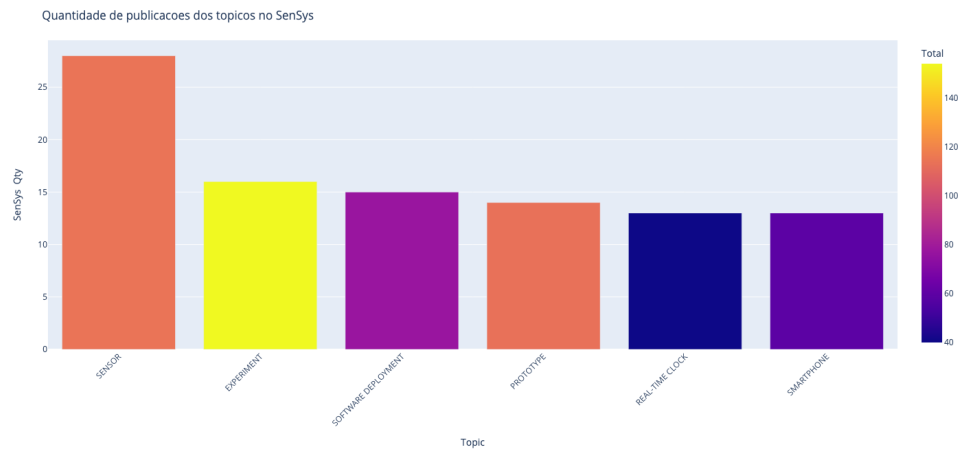
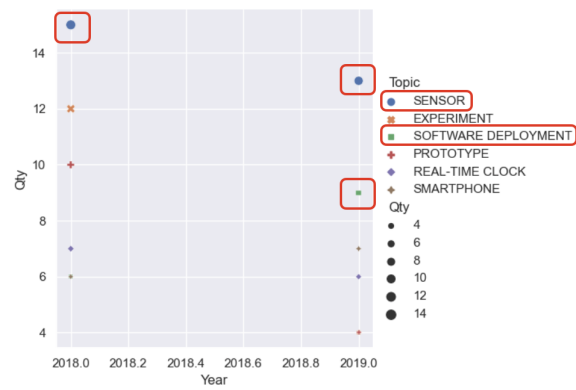


Figura 25. Tópicos mais publicados no SenSys



SenSys

Figura 26. Histograma de tópicos no SenSys

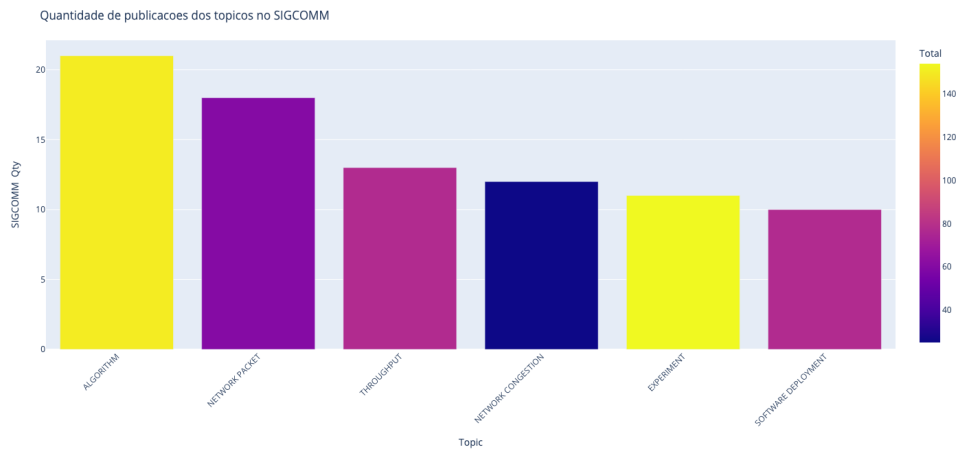


Figura 27. Tópicos mais publicados no SIGCOMM



Figura 28. Histograma de tópicos no SIGCOMM

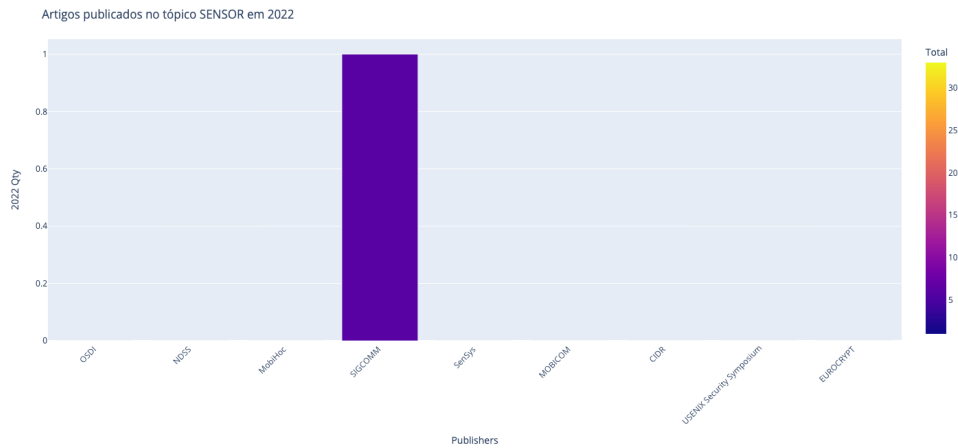


Figura 29. Artigos publicados sobre SENSOR em 2022

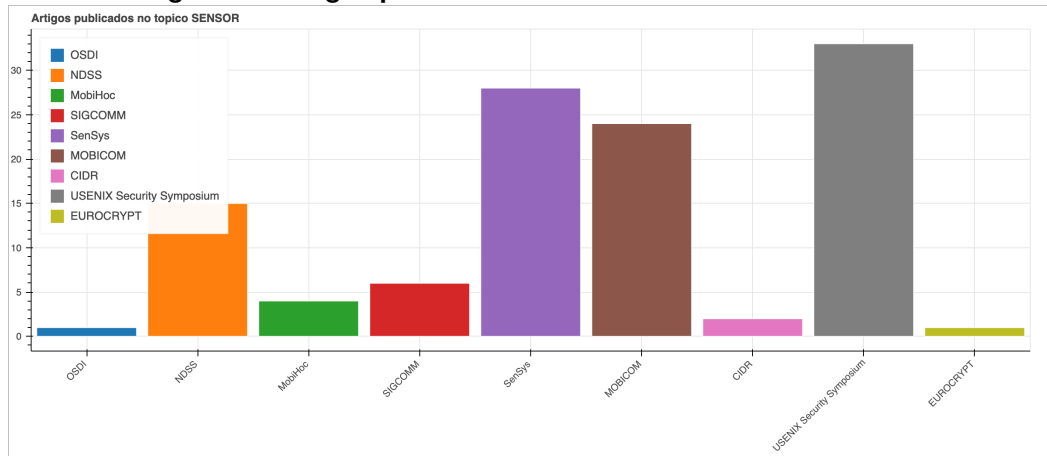


Figura 30. Artigos publicados sobre SENSOR

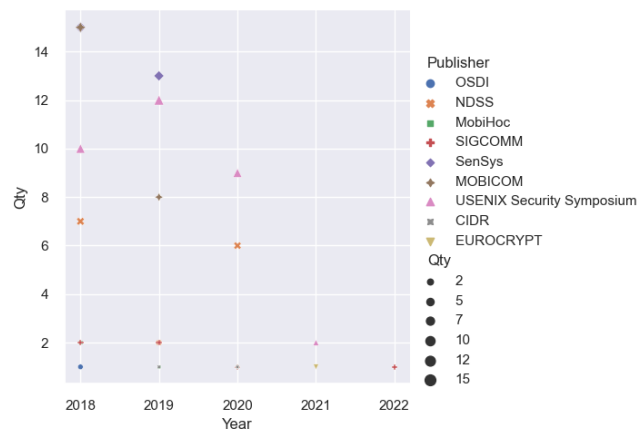


Figura 31. Comparativo hist rico dos eventos para artigos sobre SENSOR

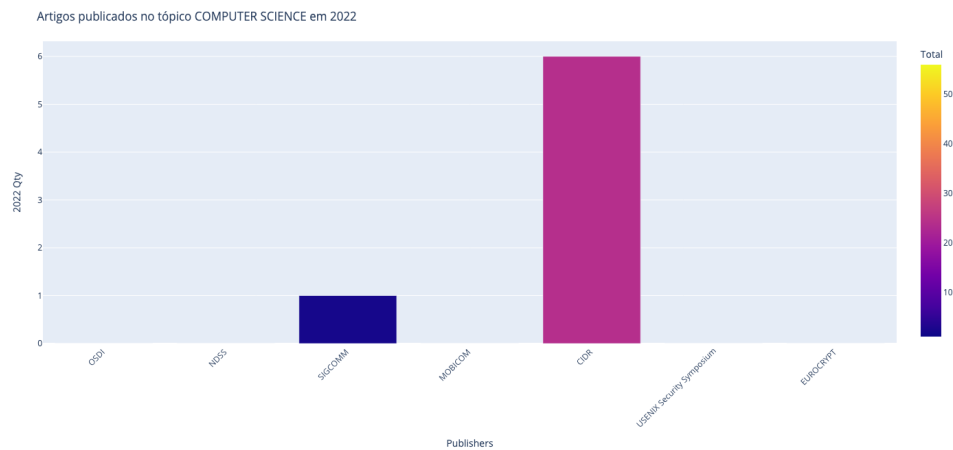


Figura 32. Artigos publicados sobre COMPUTER SCIENCE em 2022

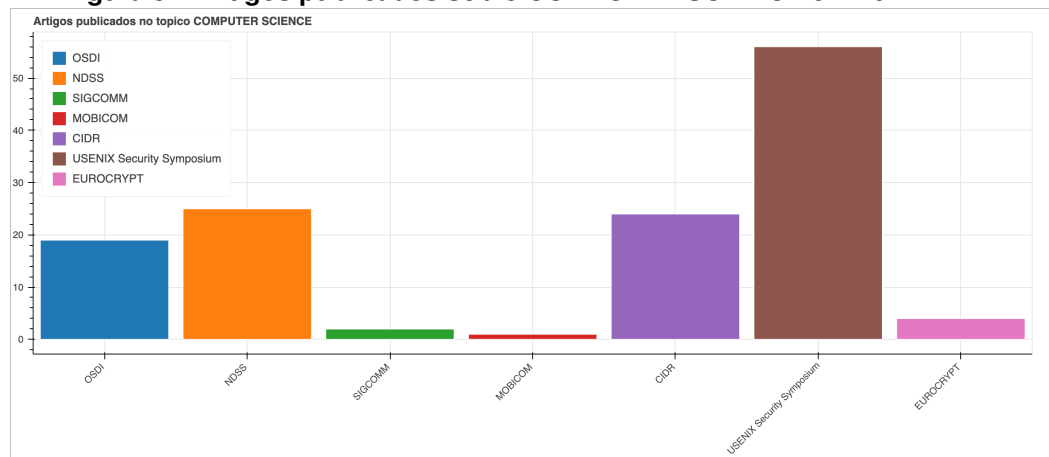


Figura 33. Artigos publicados sobre COMPUTER SCIENCE

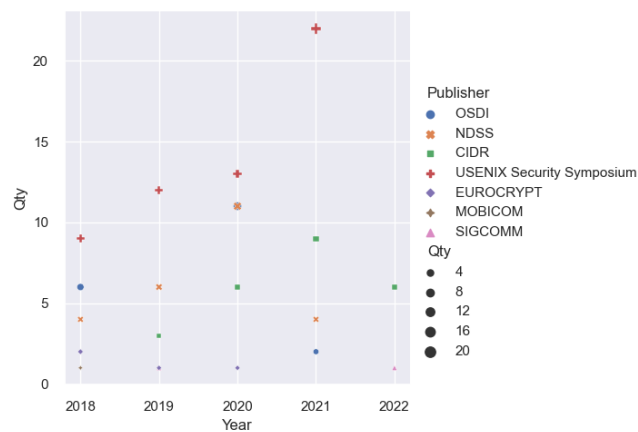


Figura 34. Comparativo hist rico dos eventos para artigos sobre COMPUTER SCIENCE

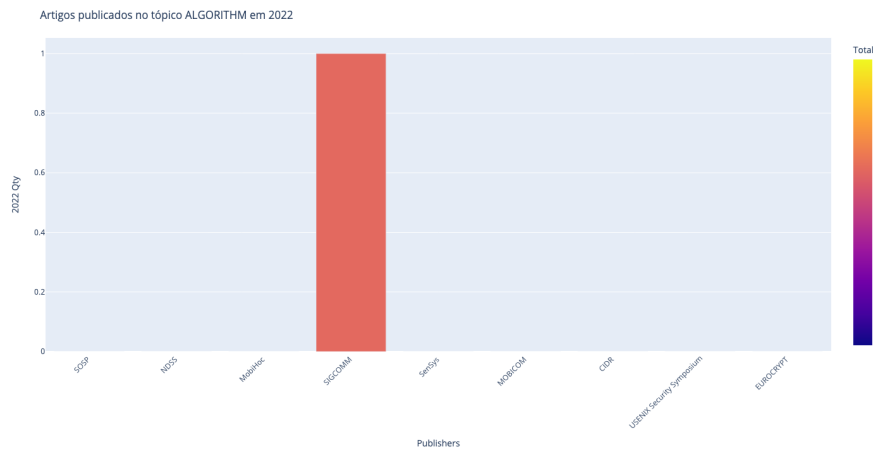


Figura 35. Artigos publicados sobre ALGORITHM em 2022

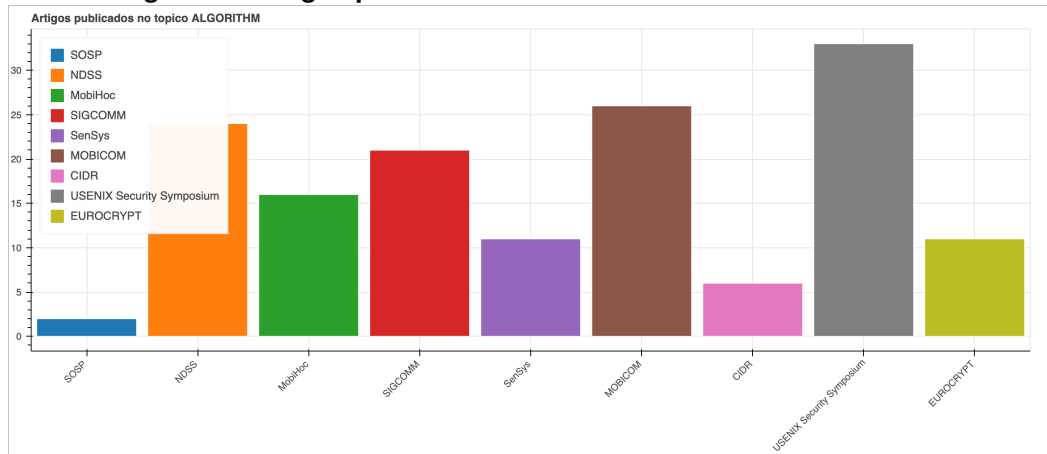


Figura 36. Artigos publicados sobre ALGORITHM

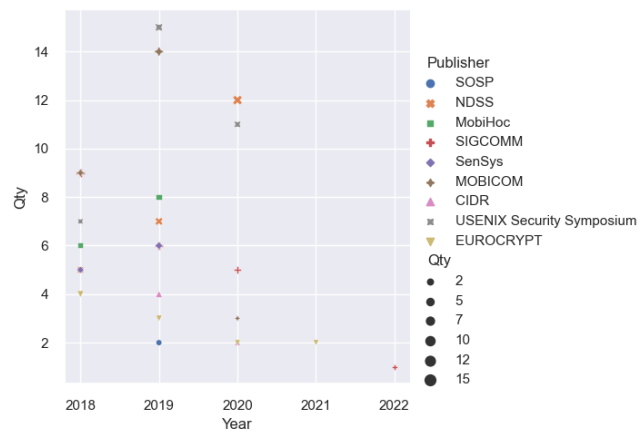


Figura 37. Comparativo hist rico dos eventos para artigos sobre ALGORITHM

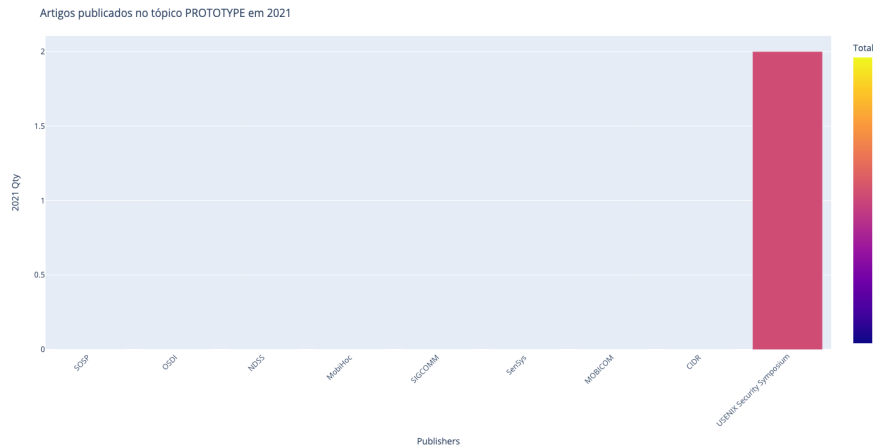


Figura 38. Artigos publicados sobre PROTOTYPE em 2022

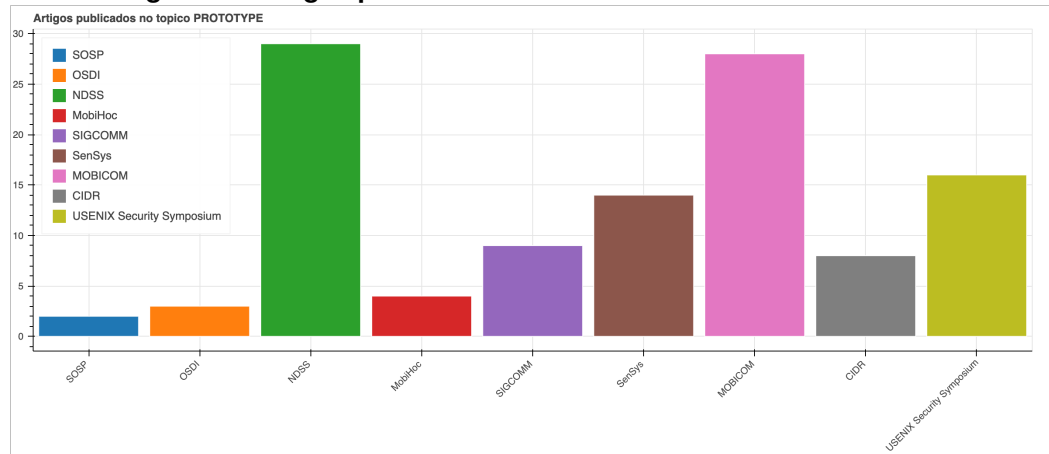


Figura 39. Artigos publicados sobre PROTOTYPE

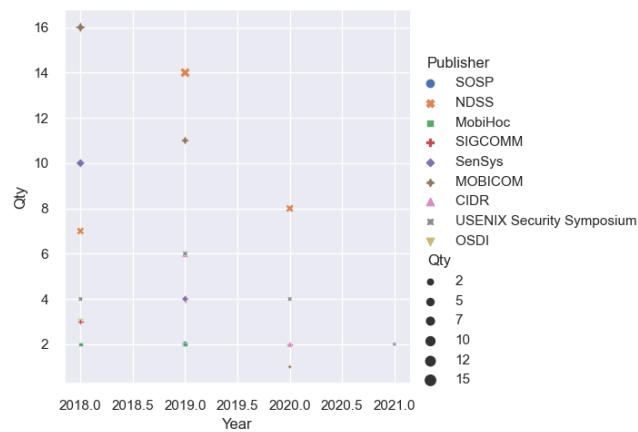


Figura 40. Comparativo hist rico dos eventos para artigos sobre PROTOTYPE

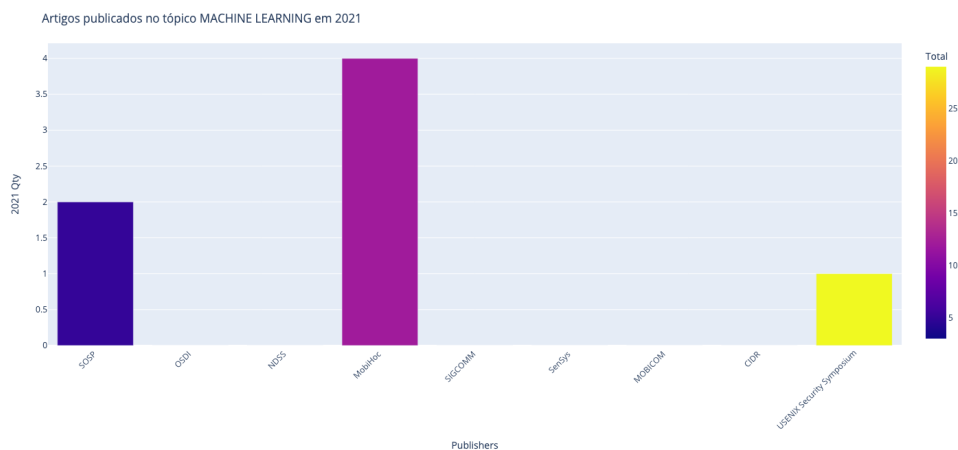


Figura 41. Artigos publicados sobre MACHINE LEARNING em 2022

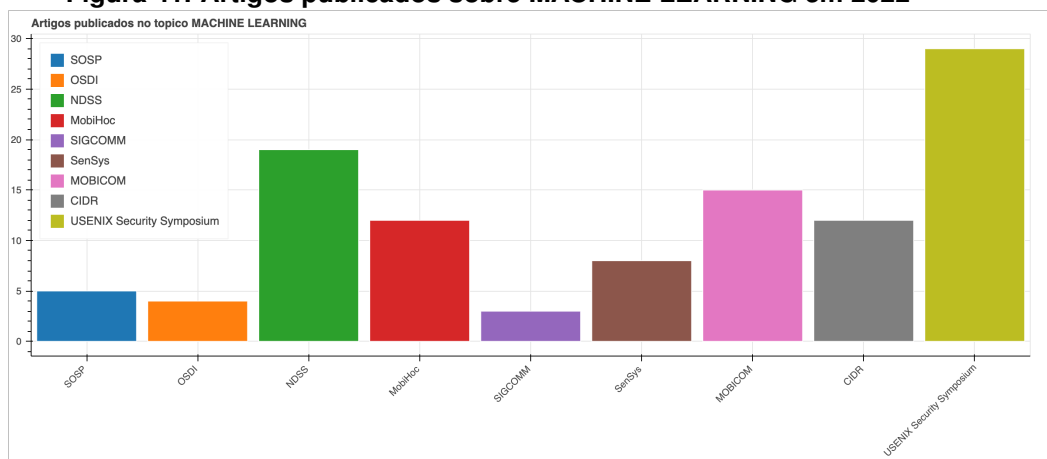


Figura 42. Artigos publicados sobre MACHINE LEARNING

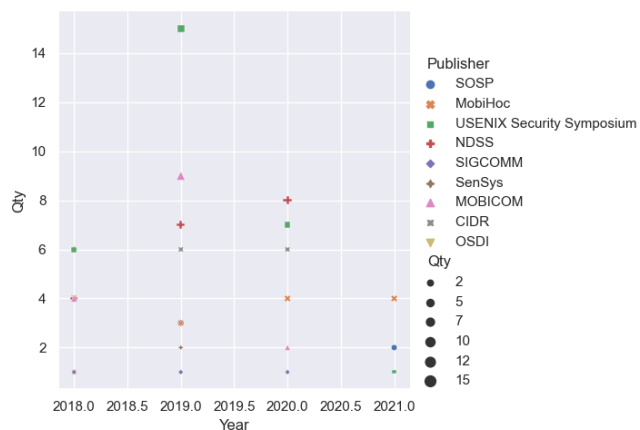


Figura 43. Comparativo hist3pico dos eventos para artigos sobre MACHINE LEARNING

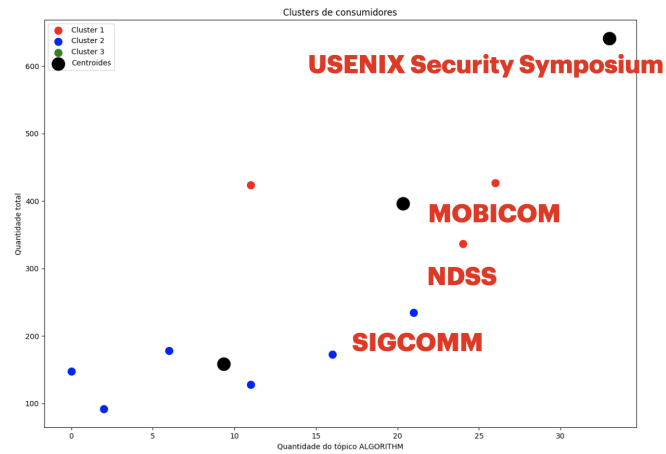


Figura 44. Clusterização do tópico ALGORITHM

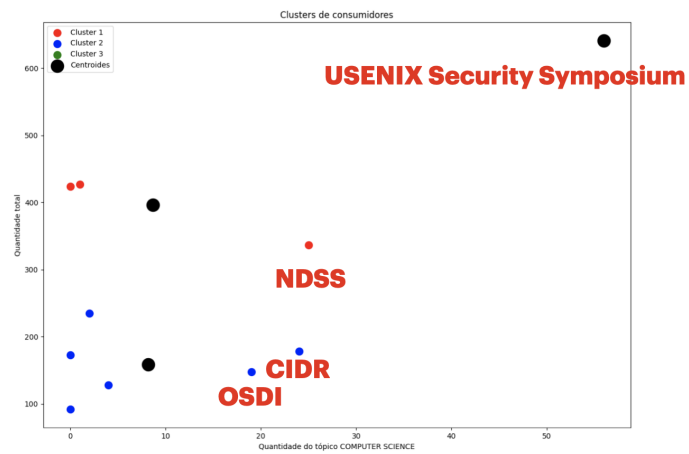


Figura 45. Clusterização do tópico COMPUTER SCIENCE

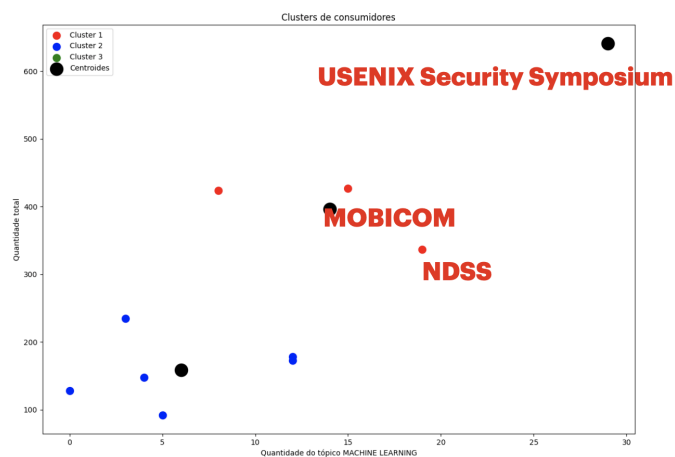


Figura 46. Clusterização do tópico MACHINE LEARNING

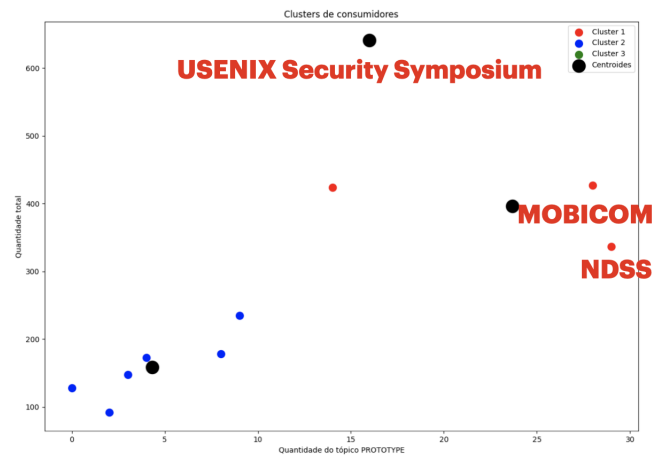


Figura 47. Clusterização do tópico PROTOTYPE

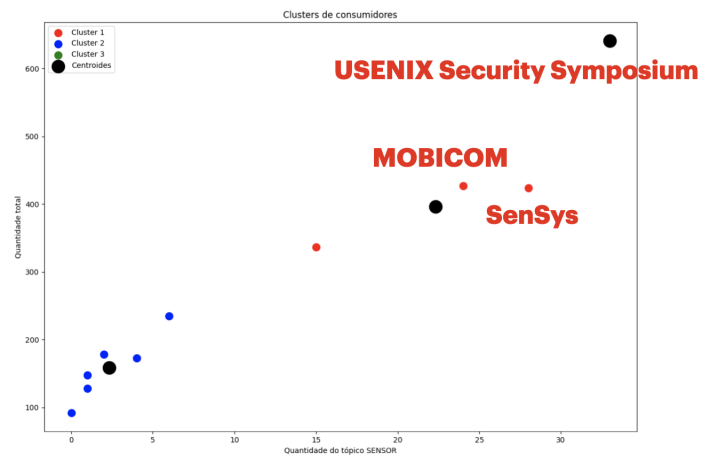


Figura 48. Clusterização do tópico SENSOR