

Elaboração de mini surveys

Seminários da EIC

Eduardo Ogasawara http://eic.cefet-rj.br/~eogasawara

Por que devo fazer uma boa revisão bibliográfica?







WWW.PHDCOMICS.COM

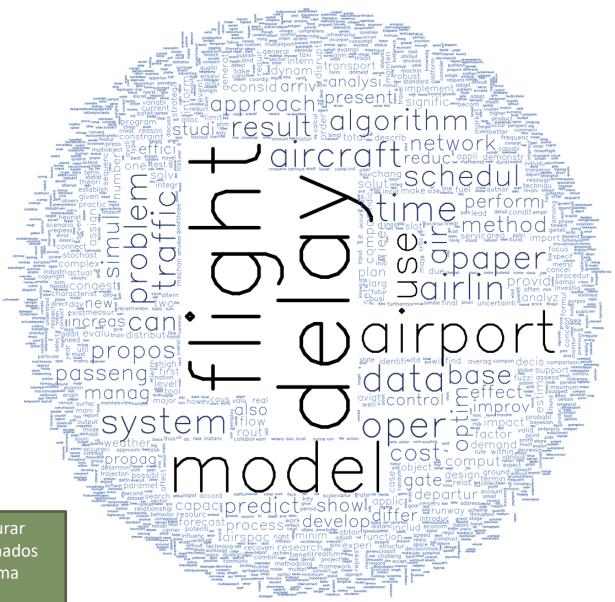
Como uma boa revisão pode ajudar?

- Motivação
 - Por que esta pesquisa é importante?
- Definição do problema
 - Qual é o problema que queremos atacar?
- Solução
 - Sua abordagem é boa, pois resolve tal determinado problema
- Quais são os trabalhos semelhantes?

Vantagens?

- Boa seção de trabalho relacionados
- Construção do referencial teórico
- Elaborar um mini-survey

Análise as tendências



Útil para estruturar trabalhos relacionados e/ou montar uma taxonomia

Taxonomia

Regarding the available literature on flight delay prediction, we have conducted a systematic mapping study. The search expression string ("airport delay" \("flight delay" \) \("predict" \\ "forecast" \\ "propagate" \)) was used to query Scopus on October 2017. Query result brought 310 references. Additionally, 29 works were added using snowballing search.

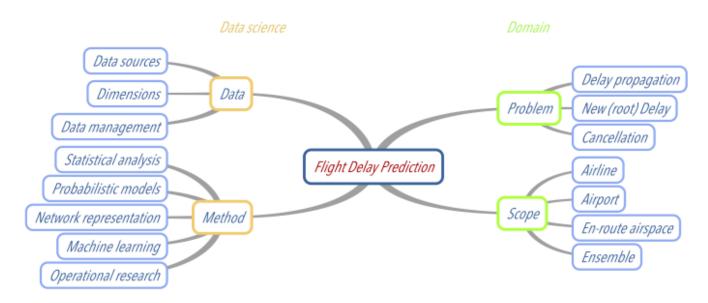
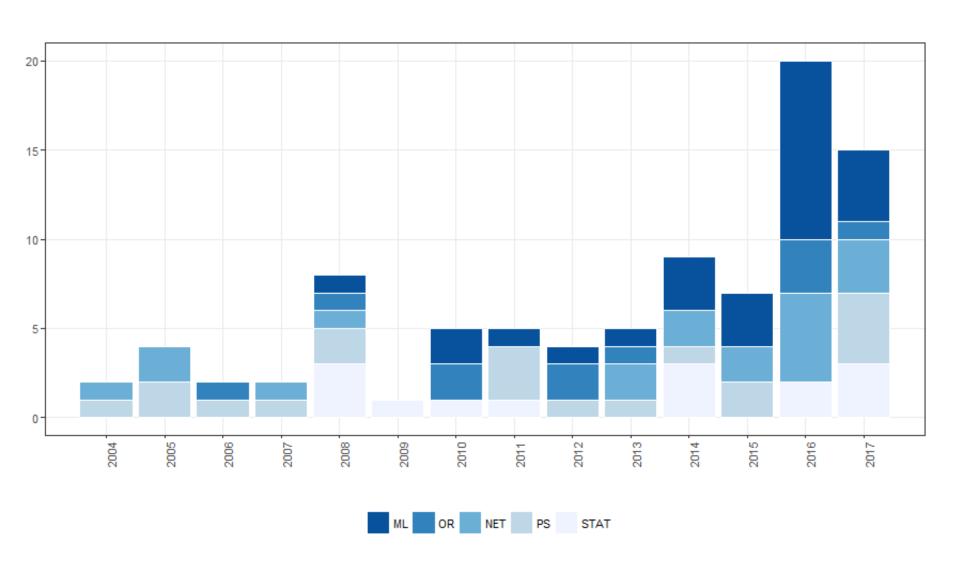


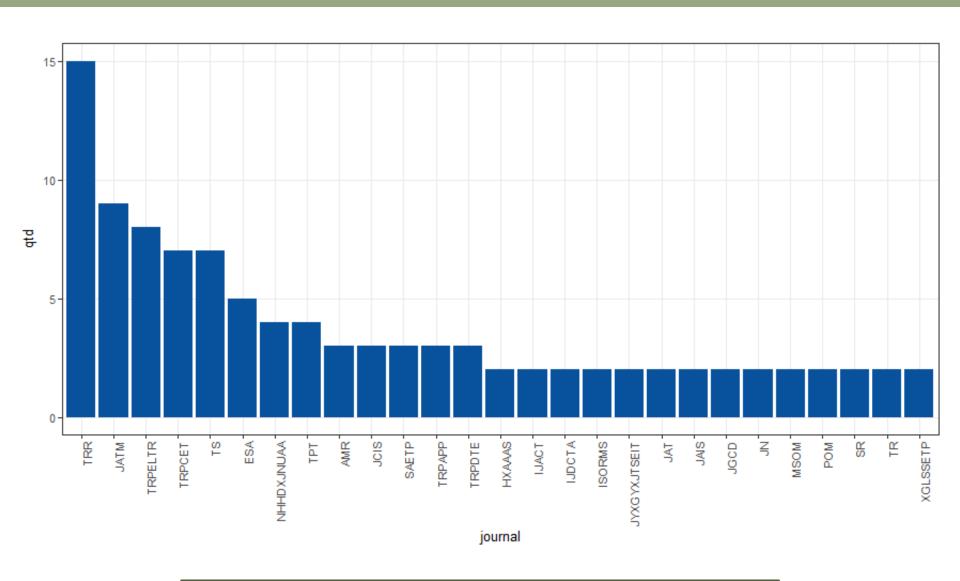
Figure 2. Taxonomy of the flight delay prediction problem

We have selected 134 to build this review due to their relevance and direct link with the flight delay prediction problem. The main criteria to be included is to have the word "delay" in the abstract, and the paper should have at least the one citation at Google Scholar per year before 2016. It means that to include a paper of 2015, it must have at least one citation, and so one.

Evolução da produção por método



Principais fóruns

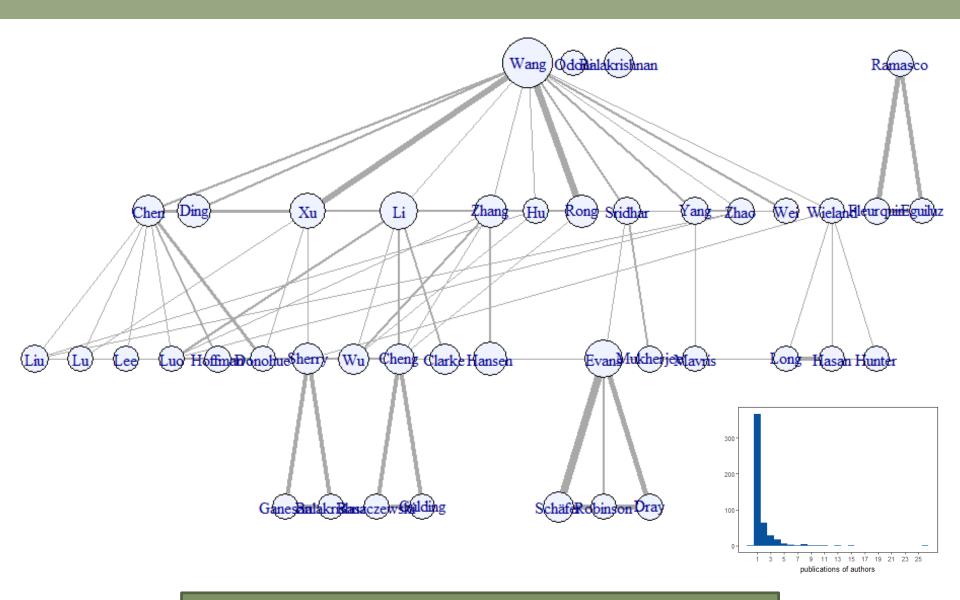


Timeline

Years	root delay or cancellation	delay propagation
1997-2000	■ Wieland (117) ■ Boswell (24)	Beatty (16)
2001-2004	■ Hansen ⁽⁵⁹⁾ ■ Mueller ⁽⁹⁰⁾ ■ Evans ⁽⁵¹⁾	Schaefer (106) ■ Abdelghany (2)
2005	■ Hsiao (62) ■ Hansen (60)	■ Wu (119) ■ Xu (123)
2006	■ Sim ⁽¹⁰⁷⁾	■ Lan ⁽⁷⁸⁾
2007	■ Wan ⁽¹¹⁴⁾ ■ Biesiada ⁽²²⁾ ■ Abdel-Aty ⁽¹⁾	
2008	 ■ Balakrishna ⁽⁹⁾ ■ Soomer ⁽¹⁰⁹⁾ ■ McCrea ⁽⁸⁵⁾ ■ Tu ⁽¹¹²⁾ ■ Pathomsiri ⁽⁹⁴⁾ 	■ Lapp (79) ■ AhmadBeygi (3)
2009	Pejovic (96)	
2010	■ Balakrishna ⁽¹⁰⁾ ■ Ganesan ⁽⁵⁶⁾ ■ Klein ⁽⁷³⁾	Ahmadbeygi ⁽⁴⁾
2011	■ Gürbüz ⁽⁵⁸⁾ ■ Evans ⁽⁴⁷⁾	Nayak (91)
2012	■ Wang $^{(116)}$ ■ Zou $^{(132)}$ ■ Azadian $^{(7)}$	■ Dück (44) ■ Wong (118)
2013	 ■ Kulkarni (76) ■ Kim (72) ■ Evans (48) ■ Xiong (122) 	■ Pyrgiotis (99) ■ Fleurquin (55)
2014	■ Rebollo (102) ■ Lin (82) ■ Baumgarten (15)	■ Campanelli (29) ■ Hao (61)
2015	■ Bloem ⁽²³⁾ ■ Cai ⁽²⁸⁾ ■ Jacquillat ⁽⁶⁵⁾	■ Ciruelos ⁽³⁵⁾ ■ Cheng ⁽³²⁾
2016	■ Choi (34) ■ Castaing (31) ■ Bertsimas (20) ■ Simaiakis (108)	■ Khanmohammadi (70) ■ Cong (36)
2017	 Takeichi (111) ■ Ding (42) ■ Baluch (13) ■ Jayam (67) ■ Jacquillat (66) ■ Pérez–Rodríguez (100) 	■ Belkoura ⁽¹⁸⁾ ■ Ben Ahmed ⁽¹⁹⁾

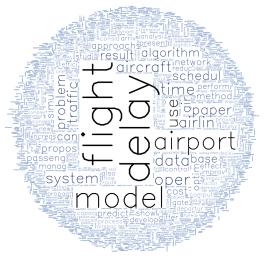
Figure 6. Time line of flight delay prediction publications: ■ Statistical Analysis, ■ Probabilistic Models, ■ Network Representation, ■ Operational Research, ■ Machine Learning

Principais autores



Outras análises

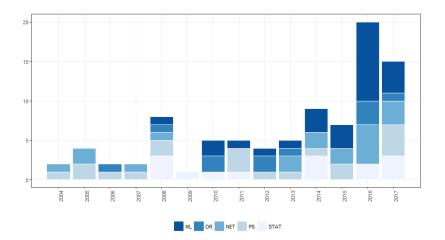
Outras visualizações/análises pragmáticas de valor

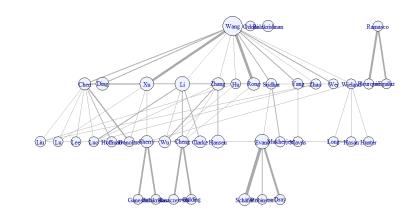


Years	root delay or cancellation	delay propagation	
1997-2000	■ Wieland (117) ■ Boswell (24)	Beatty (16)	
2001-2004	■ Hansen ⁽⁵⁹⁾ ■ Mueller ⁽⁹⁰⁾ ■ Evans ⁽⁵¹⁾	Schaefer ⁽¹⁰⁶⁾ ■ Abdelghany ⁽²⁾	
2005	■ Hsiao (62) ■ Hansen (60)	■ Wu (119) ■ Xu (123)	
2006	■ Sim ⁽¹⁰⁷⁾	■ Lan ⁽⁷⁸⁾	
2007	■ Wan ⁽¹¹⁴⁾ ■ Biesiada ⁽²²⁾ ■ Abdel-Aty ⁽¹⁾		
2008	■ Balakrishna ⁽⁹⁾ ■ Soomer ⁽¹⁰⁹⁾ ■ McCrea ⁽⁸⁵⁾ ■ Tu ⁽¹¹²⁾ ■ Pathomsiri ⁽⁹⁴⁾	■ Lapp (79) ■ AhmadBeygi (3)	
2009	Pejovic (96)		
2010	■ Balakrishna (10) ■ Ganesan (56) ■ Klein (73)	Ahmadbeygi (4)	
2011	■ G¨urb¨uz (58) ■ Evans (47)	Nayak (91)	
2012	■ Wang (116) ■ Zou (132) ■ Azadian (7)	■ Dück (44) ■ Wong (118)	
2013	■ Kulkarni ⁽⁷⁶⁾ ■ Kim ⁽⁷²⁾ ■ Evans ⁽⁴⁸⁾ ■ Xiong ⁽¹²²⁾	■ Pyrgiotis (99) ■ Fleurquin (55)	
2014	■ Rebollo (102) ■ Lin (82) ■ Baumgarten (15)	Campanelli (29) Hao (61)	
2015	■ Bloem ⁽²³⁾ ■ Cai ⁽²⁸⁾ ■ Jacquillat ⁽⁶⁵⁾	■ Ciruelos (35) ■ Cheng (32)	
2016	■ Choi (34) ■ Castaing (31) ■ Bertsimas (20) Simaiakis (108)	■ Khanmohammadi (70) ■ Cong (36)	
2017	■ Takeichi (111) ■ Ding (42) ■ Baluch (13) ■ Jayam (67) ■ Jacquillat (66) ■ Pérez–Rodríguez (100)	■ Belkoura ⁽¹⁸⁾ ■ Ben Ahmed ⁽¹⁹⁾	

Figure 6. Time line of flight delay prediction publications: Statistical Analysis, Probabilistic Models,

Network Representation, Operational Research, Machine Learning

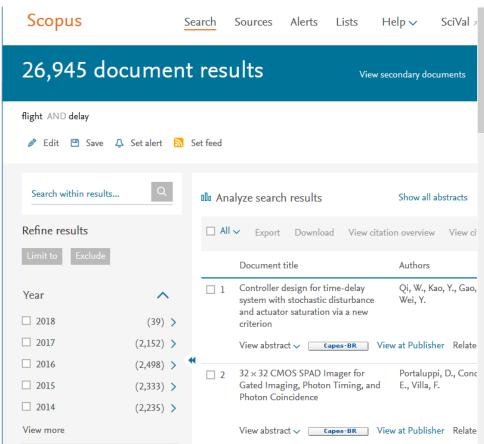




Como fazer a revisão bibliográfica? Como fazer o mapeamento sistemático?

- Suponha pesquisa sobre atrasos aéreos
 - O primeiro passo é montar uma string de busca:
 - flight delay
 - Execute a consulta no Scopus

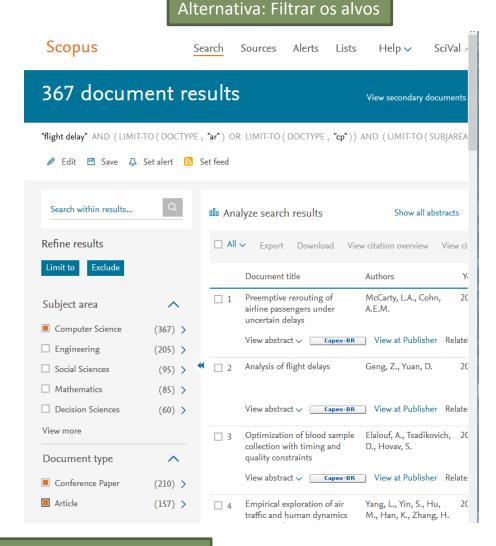




Refinando a string de busca

Melhorar a consulta

- O resultado foi muito amplo e precisa ser refinado
 - "flight delay": 1236 artigos
 - Ainda é muito
- Refinando mais
 - "flight delay" and "predict": 78 artigos
 - Incompleto
- Aumentando o escopo
 - "flight delay" AND ("predict" O R "forecast"): 268



Trabalhos relacionados: 50 a 100 artigos Mini-survey: 100 a 300 artigos

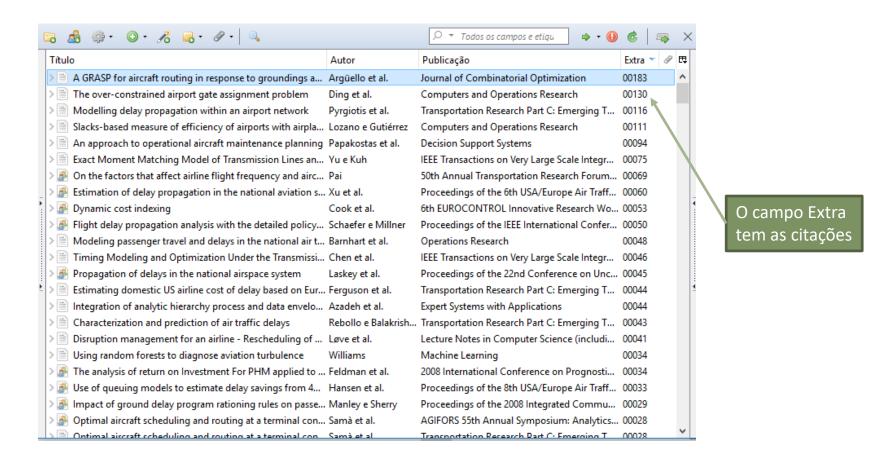
Exportar as citações

Export document settings ①							
You have chosen to export 367 documents							
Select your method of export							
Customize export							
Citation information	☐ Bibliographical information	Abstract and Keywords					
 Author(s) Document title Year Source title Volume, Issue, Pages Citation count Source and Document Type DOI 	 □ Affiliations □ Serial identifiers (e.g. ISSN) □ PubMed ID □ Publisher □ Editor(s) □ Language of Original Document □ Correspondence Address □ Abbreviated Source Title 	AbstractAuthor KeywordsIndex Keywords					

Exportar scopus.bib

Importar no Zotero

Importar no Zotero e usar a ferramenta Zotero Scholar Citations



Exportar bibtex do Zotero para importação no R

Leitura de um artigo









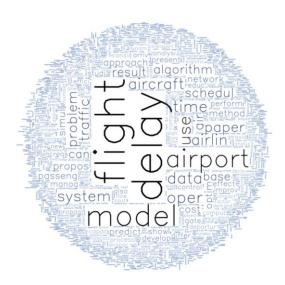
JORGE CHAM @THE STANFORD DAILY

phd.stanford.edu

Como eu leio um artigo?

- Estabeleça uma taxonomia, monte categorias de análise e critérios quantitativos para citação
- Filtre os artigos pelos critérios quantitativos
- Leia o abstract/resumo e veja se ele está relacionado ao que você pretende trabalhar
- Caso afirmativo, leia a introdução
 - Identifique a motivação, definição do problema, proposta de solução e indicativo de contribuição
 - Entra na taxonomia e nas categorias? (se não: justifique)
- Caso afirmativo, leia a conclusão
 - Cumpriu o prometido? Os resultados foram bons? (se não: justifique)
- Caso afirmativo, leia o artigo todo e faça um bom resumo
 - Anote os pontos fortes e fracos segundo uma classificação

Elaboração da taxonomia



leitura de bib text com a contabilização citations <- read.bib('scopus-count.bib') mydataset <- bibtext.dataset(citations)

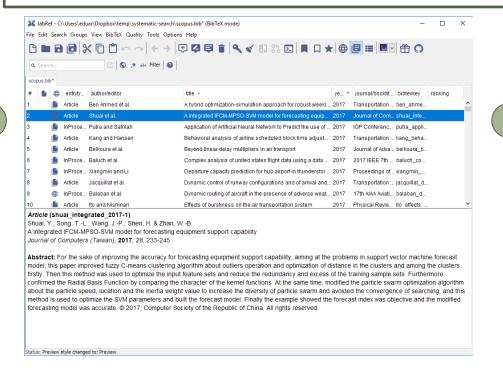
motagem de nuvem de palavras para análise myword <- bibtext.cloudword(mydataset, "publication-cloud.pdf")

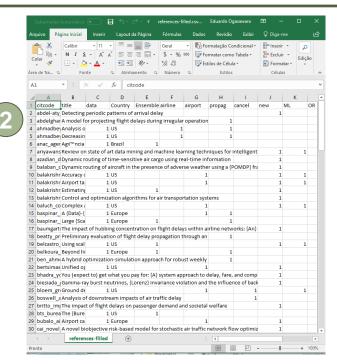
Análise de métodos / categoria

motagem de nuvem de palavras para análise (1) bibtext.saveexcel(mydataset, "references.xlsx")

#analise (leitura manual dos abstracts) e carregamento dos atributos adicionais (2) mydataset <- bibtext.mergecsv(mydataset, "references-filled.csv")

filtragem de artigos e filtro adicionais mydataset\$rate <- mydataset\$cites - (2015-mydataset\$years) > 0 mydataset\$delay <- (regexpr('delay', mydataset\$abstract) + regexpr('delay', mydataset\$title)) != -2 mydataset <- mydataset[mydataset\$rate & mydataset\$delay,]





Marcação dos artigos

- Leia os artigos completos procurando identificá-los quanto:
 - Abordagens
 - Critérios para comparação
 - Limitações
 - Diferenças em relação a sua proposta
- Procure deixar marcado nos PDF as principais características
- Prepare um parágrafo sintetizando cada artigo segundo os critérios de comparação, limitações e diferenças em relação a sua proposta

Análise de artigos para citar / não citar

```
# identifica artigos não citados ainda
citations <- read.bib('scopus-count.bib')
mydataset <- bibtext.dataset(citations)
to_add <- bibtext.diffbib(mydataset, 'references.bib')

# identifica artigos incluídos não presentes no mapa sistemático
used_citations <- read.bib('references.bib')
useddataset <- bibtext.dataset(used_citations)
snowballing <- bibtext.diffbib(useddataset, 'scopus-count.bib')
```

Citar é muito importante, mas existe a necessidade de um balanço entre as referencias citadas e o tamanho do artigo

Escrita adequada

3.3.1 Data Sources

The type of datasets from the air transportation system are mainly related to airlines, airports or ensemble. Since airlines and airports commonly do not share their databases with the entire community, they are often used by collaborators of those institutions. Ensemble datasets may include both carriers, airports, and additional information provided by governmental agencies, regulatory authorities, and service providers. Table 1 displays the type of datasets by regions. It presents the number of publications and the top three most cited papers in each category. Governmental agencies usually provide public access to their databases with different granularity. It is noticed that data from The United States Department of Transportation (43), primarily through The Federal Aviation Administration (52) and The Bureau of Transportation Statistics databases (26) are widely used to obtain information about flights. The Eurocontrol (46) database is provided by an intergovernmental organization in Europe. This dataset is also used intensively in flight delay studies (103).

Table 1

Number of sources of real data about the air transportation system per region

Region	Ensemble	Airline	Airport
Asia	2(89,111)	1 (104)	1 (121)
Brazil	$2^{(110,5)}$	0	0
Europe	7 (30,29,81)	2(109,58)	7 (103,27,96)
US	11 (90,112,128)	7 ^(78,3,4)	16 (55,10,54)

Other related datasets, such as weather, may be obtained from governmental databases or service providers. This includes, for example, The National Oceanic and Atmospheric Administration of the United States (92). In fact, authors may use more than one source to develop their models. Datasets from United States Department of Transportation (43), National Oceanic and Atmospheric Administration (92), and Weather Company (113) are commonly used to build delay prediction models.

Additionally, some researchers (130,131) create synthetic datasets to evaluate their models instead of using real data. For example, Zou et al. (131) developed a market scenario, considering airport capacity, links, frequency, and characteristics of flights and passenger demand.

Escrita menos adequada

3.4.3 Network Representation

Network representation encompasses the study of flight systems according to a graph theory. Abdelghany et al. (2) built direct acyclic graphs to model the schedule of an airline (including flight times and resources availability) to detect disruptions and their impacts on the rest of the network. They used the classical shortest path algorithm to evaluate propagation effects.

Ahmadbeygi et al. ⁽³⁾ built propagation trees to compare two different airlines, one operating in a conventional hub-and-spoke scheme and the other in a low-cost point-to-point system. Xu et al. ⁽¹²³⁾ and Wu et al. ⁽¹²⁰⁾ built a Bayesian network to model delay propagation. Baspinar ⁽¹⁴⁾ built a network-epidemic process using historical flight-track data of Europe to create a novel delay propagation model.

Dicas finais

Remover referencias não utilizadas

java -jar unusedref references.bib main.tex

Limpar campos não úteis ao artigo

java -jar latex.jar cleanbib references.bib

Mapeamento de citações (outras codificações) versus Zotero

java -jar latex.jar map-citations main.tex references-org.bib references.bib

Código no GitHub

https://github.com/eogasawara/mylibrary

Pesquisa na área

- Automatizar o processo de revisão via snowballing
- Identificação automática de taxonomia
- Identificação automática de categorias



Elaboração de mini surveys

Seminários da EIC

Eduardo Ogasawara http://eic.cefet-rj.br/~eogasawara