Australasian Transport Research Forum 2023 Proceedings
29 November – 1 December, Perth, Australia
Publication website: http://www.atrf.info

Exploring regional Queensland MaaS stakeholder perspectives through structural topic modelling

Abraham Leung[1]

[1]Cities Research Institute, Griffith University

Email for correspondence: abraham.leung@griffith.edu.au

Abstract

The research focuses on Mobility as a Service (MaaS), an emerging vision of complex socio-technical system for mobility transition, promising seamless integration at information, transaction and operational layers of transport (and also non-transport features), while allowing greater personalisation of service provision. Despite its potential, MaaS development in Australia remains in its infancy, and further understanding of the barriers and opportunities could assist the future implementation of MaaS. A structural topic model (STM) was applied to a large corpus (n=5,320) of unstructured interview text responses collected from a diverse sample of transport stakeholders (n=19). This allows for the inductive generation of a MaaS topic classification system based on three key respondent traits: stance towards MaaS, role, and location. Six topics of interest are identified, with certain topics exhibiting greater prevalence. This technique offers valuable insights for transport researchers seeking to inductively derive meaning from substantial unstructured texts, particularly in the context of emerging transport technologies. This paper showcases advancements in natural language processing to further knowledge and methodologies in transport research by presenting a STM for transport researchers aiming to analyse extensive unstructured texts inductively.

1. Introduction

In the last decade, Mobility as a Service (MaaS) emerged as a “powerful and value-laden rhetorical vision of complex socio-technical system for mobility transition” (Mladenović, 2021, p. 12) promising seamless integration at information, transaction and operational layers of transport (and also non-transport features), while allowing greater personalisation of service provision. In Queensland, MaaS is also gaining traction, in which the concept refers to a “combination of public and private transportation services accessed via an application that provides personalised journey planning, booking, and payment” (Department of Transport and Main Roads, 2023). The MaaS Program Office was created to help develop this concept and establish a government-enabled business model, and an open ecosystem MaaS. Given that Queensland is among the most decentralised state/territory in Australia in terms of population, it is essential to ensure an equitable distribution of MaaS development across the State, be it metropolitan or not. In Finland, where the idea of MaaS first emerged, though less expansive than Australia, is similarly sparsely populated. Recent stakeholder analyses have underscored the necessity of public-private partnerships (Eckhardt, 2019; Eckhardt et al., 2020) with large transport companies providing vertically integrated multi-service offerings (Fujisaki et al., 2022; Leung et al., 2023). In Japan, MaaS is also being developed, benefited by greater vertical integration from large transport companies that are also providing multi-service offerings (e.g. tourism, real estate, dining, hotels) (Hayauchi & Hidaka, 2022; Shibuya et al., 2022) – promoting tourism and equitable regional development are also main objectives of Japanese

MaaS initiatives (Ministry of Land, Infrastructure, Transport and Tourism, 2019; World Economic Forum, 2020).

For Queensland, merely replicating overseas examples may not be the best approach; successful MaaS implementation should be locally grounded that is made to fit Queensland's unique context – considering transportation planning, stakeholder dynamics, and underlying political underpinnings. This warrants an investigation of stakeholders in Queensland to better understand their perceptions of MaaS.

2. Method

2.1. Data collection

This study aimed to explore the potential of unsupervised machine learning textuak analysis to a stakeholder engagement exercise sought to understand the potential of MaaS in at least three regional cities/towns within at least two different TMR regions (Leung et al., 2021). For MaaS to be considered a viable option in a regional context in Queensland, specific conditions must be met. These hypothesised conditions include:

1. A minimum population of 10,000 (rural or sparsely populated locations are not considered in this study);
2. The existence of regular public transport services that could be enhanced by MaaS;
3. A transient workforce (e.g., agricultural, mining or defence force employees) who may not own personal vehicles or additional personal vehicles; and/or
4. A considerable demand for mobility from tourism (e.g., a transient population without personal vehicles)

Based on these criteria, three regional case study sites were selected:

1. Townsville
2. Rockhampton (including Yeppoon)
3. Gladstone

Stakeholder engagement was conducted in the three selected study areas to comprehend the existing transport challenges and evaluate MaaS as an opportunity for regional areas. To obtain more in-depth insights, the majority of the engagement involved one-on-one interviews. Due to COVID travel restrictions, most interviews were carried out using online video conferencing platforms (e.g., MS Teams or Zoom). At the request of a specific agency, one workshop was organised with multiple participants at the same time.

As shown in Table 1, nineteen stakeholders (three women and sixteen men) were recruited to participate in twelve research interviews and one workshop (with five participants). The participants included State Government (including Translink), local governments, various transport service providers, key transport users/destinations (e.g. universities) and transport academics/professionals who are well-versed in MaaS. A majority (seventeen) of the participants were across Townsville, Rockhampton (including Yeppoon) and Gladstone. In addition to regional Queensland stakeholders, two external experts from New South Wales and the United Kingdom, respectively, were interviewed. The total size of the sample, including external experts, is n=19. One pilot interview was not included to ensure data quality.

Table 1: Participant profile and location (n=19)

Type of participant	Townsville	Rockhampton (incl. Yeppoon)	Gladstone	Across Queensland	Outside Study Regions
State Government (Queensland)	n/a	n/a	n/a	(n=3)	n/a
Local Governments	Townsville City Council (n=5)	Rockhampton Regional Council* (n=1)	Gladstone Regional Council (n=1)	n/a	n/a
Transport providers	Bus (n=1)	Bus (n=1)	Bus (n=1)	Point-to-point (n=2)	DRT (n=1)
Users, academic or experts	n/a	University (n=1)	n/a	n/a	NSW, n=1) UK (n=1)

* Despite multiple requests, no interview was able to be conducted for the local government area of Yeppoon, which is under the jurisdiction of Livingstone Shire Council.

The interviews and the workshop were generally of 1-hour duration, in which they explored the following key questions about MaaS:

- Knowledge and views about MaaS
- Unique mobility needs in the study area regions
- Opportunities for MaaS (both for communities and transport service providers)
- Enablers and barriers to implementing MaaS in the study area regions
- View about user acceptance
- Collaboration
- Suggestions and recommendations

The guide questions were open-ended, specifically designed to address the primary research question concerning stakeholders' views on MaaS in regional Queensland. At least one experienced moderator facilitated each interview and workshop session, guiding discussions to ensure all participants contributed responses and asking probing questions when necessary. Interviews were recorded and transcribed to facilitate subsequent textual analysis. The researchers transcribed the focus group recordings verbatim. The study protocol received approval from the Griffith University Research Ethics Committee (GU ref no. 2020/579). To ensure anonymity, respondents were deidentified and grouped during the analysis process.

2.1. Topic modelling and automated content analysis

Topic models are unsupervised techniques utilised for automatically organising, understanding, searching, and summarising text documents (Blei, 2012). Topic modelling is a computer-assisted content analysis process in which a collection of texts is coded into substantively meaningful themes, known as 'topics' These topics are not predefined but instead emerge inductively as algorithms discern the hidden patterns underlying a set of texts. Such methods are generally referred to as automatic content analysis (ACA), with earlier models originating in the 1990s (Deerwester et al., 1990). However, it was not until the 2000s, thanks to increased computational power and the development of more advanced techniques, that machine learning of content truly matured (Nunez-Mir et al., 2016).

One particularly influential method developed after the turn of the millennium is the latent Dirichlet allocation (LDA) approach, introduced by Blei et al. (2003). LDA is a three-level hierarchical Bayesian model designed to iteratively infer the concepts present in a text corpus and determine the proportion of the literature in which they occur. This approach represents a significant advancement over previous models, as it abandons the ‘bag-of-words’ assumption – the notion that word order within a document is irrelevant – and accounts for the exchangeability or interchangeability of specific words. Consequently, this method enables unsupervised concept seeding that is more objective and less susceptible to human input and bias.

The structural topic modelling (STM) approach, proposed by Roberts et al. (2013), is regarded as a successor to LDA. STM incorporates covariates to accommodate metadata associated with documents in a corpus, allowing researchers to model not only topical content but also the relationships between topics and various document-level covariates. Here, metadata refers to the underlying variables (data of data) associated with the unit of study. This distinctive feature of STM enables a more refined understanding of thematic patterns within the data and the factors influencing topic formation. Initially designed for analysing political texts, STM has been widely adopted in social sciences (Roberts et al., 2016). However, its application in transport is limited, except in some instances in literature reviews (Bai et al., 2021; Das et al., 2017; Tamakloe et al., 2022), incident records or survey responses (Bardutz & Bigazzi, 2022), and text mining analysis (Kuhn, 2018). This paper aims to apply STM to interview data analysis in MaaS research for the first time.

Topic modelling, an unsupervised text analysis method, identifies, categorises, and extracts information by clustering frequently co-occurring words across a collection of documents to discern different and prevalent topics (Blei, 2012). This computer-assisted text analytical technique offers several advantages over the traditional human subjective coding procedure in focus group analysis, including mitigating observer dependency bias, faster processing of large volumes of text, and consistent treatment of all documents.

For this dataset, the collection of documents comprised transcripts from the interviews and focus groups. Each document contained text transcriptions of discussions (i.e., all verbal exchanges among participants and the moderator).

2.2. Data analysis

Data processing and analysis were conducted using the STM package in R (Roberts et al., 2019) to obtain the following results: (i) a set of topics, (ii) a set of keywords representing each topic, (iii) the prevalence/expected proportion of each topic in the collection, and (iv) the covariate estimates using three metadata fields:

- Respondents’ stance to MaaS (enthusiastic or reserved),
- Respondents’ role (government or not); and
- Respondents’ location (whether they live in the same area of work (locally-based)).

To apply STM, the transcripts underwent data pre-processing as follows:

(1) Introduction and greetings at the beginning of each interview were removed, as well as the introductory sentences under each guide question.
(2) Common stop words, such as ‘a’, ‘the’ and ‘we’ which have limited semantic value, were removed by default settings.

(3) Customised stop words, such as ‘can’, ‘yes’, ‘also’, ‘stuff’, and ‘really’ were removed.
(4) Words with different tenses or forms were consolidated to their word stem, for example, ‘commerci’ for ‘commercialise’, ‘commercial’, and ‘commerce’

Before processing, there were 2,321 unique terms (6,513 tokens or pre-processed words). These were reduced to 1,128 terms and 5,320 tokens. The unit of study, ‘document’ was composed of text transcriptions of a certain respondent. In total, 19 documents entered the study, corresponding to the number of valid respondents. While it is also possible to use the questions asked as the unit of study, metadata would become multi-dimensional and difficult to interpret, and hence this is not considered in this exercise.

These pre-processing steps ensured the data was in an appropriate format for STM analysis. By reducing the complexity of the dataset and focusing on relevant terms, the subsequent modelling was able to identify meaningful topics more effectively. The use of metadata fields, such as the respondents’ stance on MaaS, role, and location, allowed for a more in-depth understanding of the relationships between topics and various background covariates.

In addition to the main topic words, additional analysis were conducted to explore the relationship between topics and document-level covariates such as the stance towards MaaS, respondent role, and location. To do so, an estimation model was fitted to the data based on the following vector covariate model as in Roberts et al. (2019):

$$Prevalence_{ij} \sim \beta_0 + \beta_1 \times Stance + \beta_2 \times Role + \beta_3 \times Location + \varepsilon_i$$

Here, i indexes the i^{th} stakeholder respondent and j indexes the jth topic. $Prevalence_{ij}$ is the matrix of topic prevalence for each respondent's interview transcript obtained from conducting the topic model analysis. Stance is a continuous variable ranging from 1 to 3, gauged by three questions on the respondents' views and opinions of MaaS. Role and Location are categorical variables. This additional analysis can provide insights into how different factors may influence topic formation and the distribution of topics across the dataset, enabling a deeper understanding of the context in which the topics emerge.

2.3. Topics number determination

To apply STM effectively, the analyst must establish the number of topics beforehand, similar to what is performed in clustering. The determination of topics is typically achieved by examining the dataset and calculating four key metrics: 1) held-out likelihood, 2) semantic coherence, 3) residual, and 4) lower bound. Semantic coherence evaluates the co-occurrence of words within the documents, ensuring that the selected keywords pertain to a single concept, thus preserving the topic’s interpretability or quality (Mimno et al., 2011). Held-out likelihood estimates the probability of keywords appearing in documents to indicate the topic model’s generalisation capability (Wallach et al., 2009). The residual and lower bound metrics provide additional information about goodness of fit and the whether overfitting occured, respectively. Further information of the topic number determination process can be found in the *stm* R package guidelines (Roberts et al., 2019). These metrics help evaluate the quality, interpretability, and predictive accuracy of topics generated by the model. The four metrics were computed for a range of 3–20 topics to identify the model that yields the most semantically coherent and distinct topics. At the end, a six-topic model was selected, as it offered an optimal balance of relatively high semantic coherence and held-out likelihood while maintaining low residual and lower bound values (Figure 1).

Figure 1: Model diagnostics to determine the number of topics

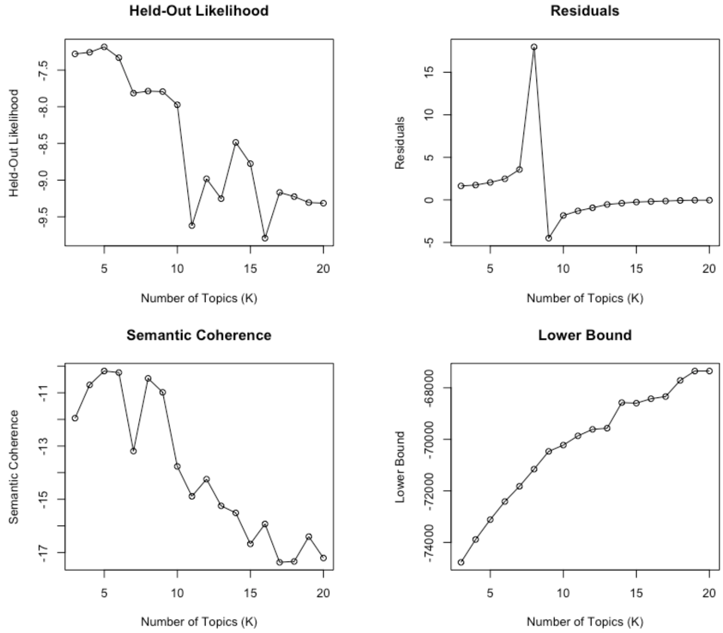
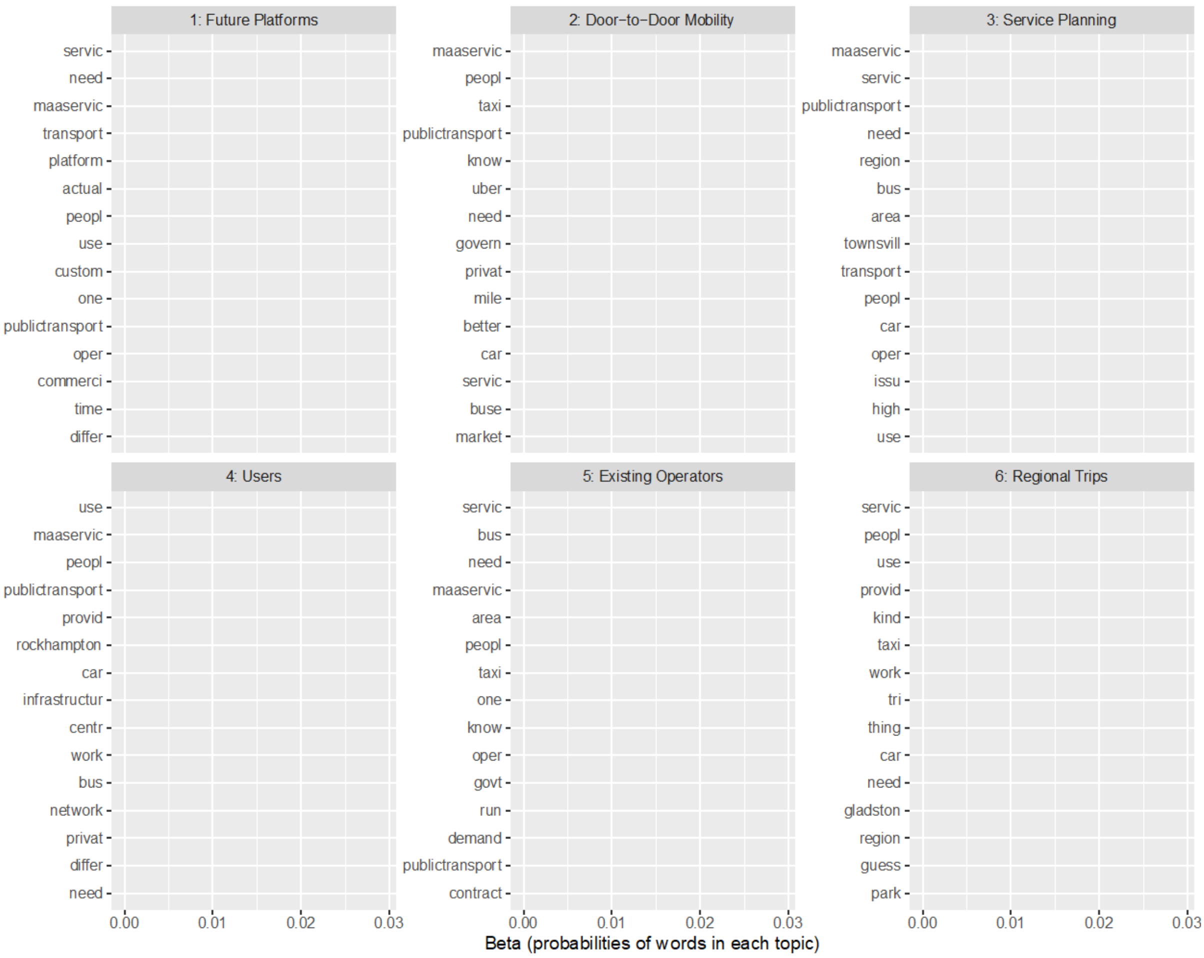


Figure 2: Highest word probabilities for each topic



3. Results and discussion

3.1. Topic analysis results

The same textual dataset was previously featured in a project report (Leung et al., 2021) and a subsequent publications (Leung et al., 2023; Leung & Burke, 2021). The current study aims to explore the usefulness of a computer-driven STM approach, as well as provide a more comprehensive understanding of the topics and their relationships with various document-level covariates.

A topic label was assigned to each topic, based on the word terms contained within. These topics are discussed further in the following section. Table 2 displays the prevalence and the most representative words for each topic, based on the highest probability and FREX metrics. The probability indicates the occurrence of a term within a given topic, while FREX (FRequency and EXclusivity) is calculated according to the word's frequency and its degree of exclusivity to a specific topic (Airoldi & Bischof, 2016). FREX attempts to find words which are both frequent in and exclusive to a topic of interest. Here, exclusivity measures the extent to which the top words for each topic do not appear as top words in other topics. Figure 2 presents the 15 words with the highest probability in each topic that assists the interpretation of the topic and label assignment. Topic proportion refers to the share of the terms contained in each topic.

Table 2: Description of the topics identified – respondents (n=19)

Topic label	Topic proportion	Metric	Word term 1	Word term 2	Word term 3	Word term 4	Word term 5	Word term 6	Word term 7
3: Service Planning	0.28	Prob.[#]	maaservic	servic	publictransport	need	region	bus	area
		FREX	townsville	lack	social	high	drt	island	disadvantag
5: Existing Operators	0.22	Prob.	servic	bus	need	maaservic	area	peopl	taxi
		FREX	demand	hour	run	school	contract	ondemand	know
1: Future Platforms	0.14	Prob.	servic	need	maaservic	transport	platform	actual	peopl
		FREX	platform	qualiti	commerci	actual	player	avail	often
6: Regional Trips	0.14	Prob.	servic	peopl	use	provid	kind	taxi	work
		FREX	kind	guess	gladstone	event	trial	budget	facil
4: Users	0.13	Prob.	use	maaservic	peopl	publictransport	provid	rockhampton	car
		FREX	rockhampton	infrastructur	network	centr	conveni	activ	cbd
2: Door-to-Door Mobility	0.09	Prob.	maaservic	peopl	taxi	publictransport	know	uber	need
		FREX	mile	success	train	know	abl	taxi	first

[#]Prob. = Highest probability

3.2. Topic results and key quotations

The topical results derived from the STM analysis are summarised in the subsequent section. As anticipated, the terms 'MaaS' or 'services' appeared frequently, since they pertain to the central topic of this research. The researcher attempted to remove these common terms to test their impact on the model; however, their exclusion resulted in less satisfactory model convergence. Consequently, these terms were retained in the final STM run. Some example narratives of each topic are listed below. STM has been used for narrative/quotation identification, especially very large datasets. The experience of automated (this paper) and manual identification (Leung et al., 2023) of the same interview data can be compared. The six machine-generated topics encompasses a range of issues, from *Service Planning* and *Existing Operators* to *Future Platforms*, *Regional Trips*, *User*s, and *Door-to-Door Mobility*. The following part outlines these topics in more detail:

Topic 3: *Service Planning* emerged as the most prevalent topic, accounting for 28% of the terms. These entries were predominantly contributed by government stakeholders, including both State and local government representatives. A higher proportion of the discussion on this topic originated from Townsville, with some mention of potential tourism applications of MaaS and on-demand services, as illustrated by the selected quotes below:

> *'Public transport is a basic social responsibility. I do not expect that MaaS may suddenly transform services into massively productive routes – it is going to help, but I do not think it is going to be that transformative' (State government respondent)*

> *'I think (MaaS) could really assist with people who are coming here from overseas, or even from interstate. If they don't have a car, MaaS can definitely assist them in getting around, beyond the public transport network, taxies and rideshare. Maybe there could be a tourism kind of private services, or airport shuttle that could be part of that same ecosystem and (users) without going on Google to looking up every transport provider in the region and not sure what's out there. I think a big benefit of MaaS is that it will reduce the transfer penalty on the information side and also the connection side. Potentially, if you can get other operators on board, they can provide the services.' (State government respondent).*

> *'MaaS is not a very common term, and the understanding needs to be localised and popularised - and cater it for the individual needs. We need to develop services by knowing what do they value? A key one is simplicity and convenience.' (Local government respondent, Townsville.*

This substantial prevalence of *Service Planning* as a topic emphasises the importance placed on planning and implementation of MaaS in regional Queensland. The involvement of government stakeholders from different levels highlights the need for collaboration and coordination in addressing the challenges and opportunities presented by MaaS. Furthermore, the inclusion of tourism applications and on-demand services underscores the potential for MaaS to cater to diverse transportation needs in regional areas, thus enhancing the overall user experience.

Topic 5: *Existing Operators*, accounting for 22% of the discussions, emerges as the second-largest topic. The majority of contributions in this topic come from transport operator respondents, representing both public and community transport sectors. Key terms include 'service' 'need,' 'demand' and 'contracts':

> *'It think it should be same app for each city in Queensland but you have to be aware with where you rolled it out. (You) need to do a lot of market testing with the app, pricing, how would people use it – even if it is a same price now but all in one place. And would you use it if it is a subscription model, etc.' (Public transport service provider)*

> *'Well, obviously to provide this service you want to be reimbursed for. And that's where I guess in regional centres on-demand gets a bit interesting because you don't have the high density housing. And let's be honest, on-demand buses in country areas are probably borderline taxis, but in saying that, whilst I'm obviously a bus operator and we want to naturally grow like any other operator. You've got to ask the question in a regional area: Should taxis be trying to grown into MaaS more than buses grown into MaaS?' (Public transport service provider)*

'MaaS would benefit us if we were to become integrated into the entire public transport network – because there's an awful lot of assets that are sitting around that are not deployed that are just chewing up organisation's money..... aggregate services to make best use of vehicles and save cost of clients.' (Community Transport service provider)

Topic 1: *Future Platforms* holds the third-largest share at 14%. This topic is primarily contributed by respondents with an interest in MaaS development or research. Frequent terms within this topic include 'platform', 'service', 'need' and commercial aspects.

'So I think in order for MaaS to get off the ground, there needs to be quality services that can be incorporated up into the platform. The platform needs to allow customers to pay using a variety of method' (Private transport service provider).

'It is important to understand users' travel needs, which then we can develop designs to adapt MaaS platforms to meet such needs' (MaaS developer)

'You need to have the different mobility services – if you don't have different players in the market, there is nothing to integrate… one thing is just better and the ability to coordinate different assets to use it in a way that is available to people, not only the physical availability, but the informational availability as well. So if we can do that through MaaS and other integration initiatives and organising integration if you can pull different community assets together. I think a one-stop-shop kind of manner. "The whole will be more than the sum of its parts". You can enhance the service offering to a lot of people.' (MaaS and public transport expert)

In fourth place is Topic 6: *Regional Trips*, representing 14% of the discussions. Most contributions to this topic come from Gladstone, a mining town with a higher prevalence of regional and long-distance work-related travel. The terms in this topic reflect a greater focus on industrial employers. However, some respondents from Townsville also contribute to this topic, discussing the relative isolation of islands or remote areas that are also tourist destinations.

'A lot of them (industry employers in Gladstone) do have their own bus services to do it. So they don't have to provide mass parking on their site. They would shuttle people in from the various regions to the terminals that they were working at. And they incentivised their employees so if they did get knocked off early it is all financially paid to them if they took the bus as opposed to driving their own cars.' (Local government respondent, Gladstone)

'Public transport currently does not serve major recreation or entertainment facilities in town. Magnetic Island – Services could not meet demand' (Local government respondent, Townsville)

'There is free parking (at regional cities), and most people drive. Current bus service levels do not support journey to work as adequately as they could.' (State government respondent)

Topic 4: *Users* occupies the fifth position with a 13% share. This topic predominantly features contributions from local government stakeholders expressing user needs and challenges in transportation. Major concerns include unmet local community transport needs and the lack of active travel infrastructure in the hotter climates of Queensland's regional areas, which is also due to lower population densities. This issue may hinder the development of MaaS.

'Public transport is a bit hard, but we are trying to work in that space to increase at least more on walking and cycling (active transport) side of things to increase mode use and patronage in those areas. There were discussions about how to price headwork infrastructure to encourage infill development. Having more dense living might support MaaS, but it could be a long process.' (Local government respondent, Rockhampton)

'Townsville is a hot city for active transport – even with MaaS you still need to walk some of the way. For active modes to work needs pathways with more trees or shade to make it more comfortable to get around, be mindful that there is heat also during nights.' (Local government respondent, Townsville)

Lastly, Topic 2: *Door-to-Door Mobility*, the smallest topic, accounts for 9% of the discussions. This topic refers to the (lack of) connectivity between different modes of transportation and the desire for point-to-point travel in a regional context. Contributions to this topic are more likely to come from local governments, users, and the private ride-booking industry.

'People now can already do multimodal travel, just it is over different platforms, for payment or booking. I think in the case of MaaS, if the substantive part of the journey is going to occur on a mass transit vehicle, so they do the heavy lifting as it were, and they do the large kilometres and we're talking about things like taxis, Ubers, scooters, bike shares to do the last kilometre. However, the market operates in silos – the different modes "do not talk to each other", MaaS could force operators to think about the consumer – it starts to think about what actually starts to shape thinking around the way that consumers see service and start to think about whole of journey solutions for consumers.' (Private transport services provider)

'It is possible to have more competition in regional areas, just there not enough market. Still Uber and taxies do complete (in the ridebooking market). Subsidies distort markets, but the bus which takes feeder customers won't exist without subsidies' (Rockhampton user group).

In summary, the analysis of the six identified topics provides a comprehensive understanding of stakeholder perspectives and concerns related to MaaS implementation in regional Queensland. The STM approach were more time-efficient compared to traditional manual or semi-computer aided approaches. However, the quotation identification still require some degree of manual inspection and understanding of the original text. This computer-driven approach allows for a rapid exploration of the topics associated with MaaS in the study area.

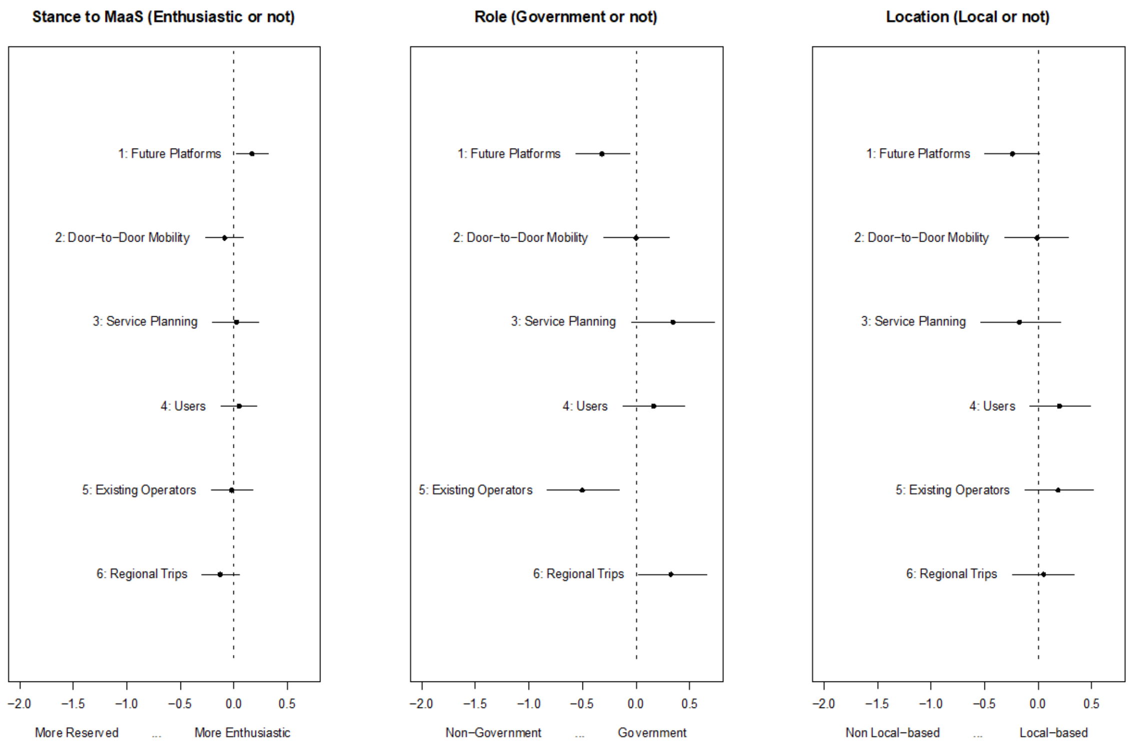
3.2. Effect of respondent characteristics on topics

A key advantage of STM is its ability to establish relationships between topic prevalence and metadata rapidly. Metadata in this study refers to the information about a respondent's background and characteristics, as well as their stance on MaaS (assessed through probing questions on their understanding and preferences).

Figure 3 provide a graphical depiction of the three metadata marginal effects on the six topics. The theme *Future Platforms* was most likely to be mentioned by individuals who were more enthusiastic about MaaS. In contrast, those with more reserved views of MaaS tended to focus on *Regional Trips* and *Door-to-Door Mobility* topics. Regarding the role of respondents, government representatives placed greater emphasis on *Service Planning*, *Regional Trips*, and *User* topics. Conversely, topical issues concerning *Existing Operations* and *Future Platforms* were mostly contributed by non-government stakeholders.

Given the decentralised population and settlement pattern of Queensland, often transport policies/practice were made remotely (e.g. State government officers may make decisions to regional areas whilst usually based in Brisbane, the state capital). The metadata of "locally-based" or not tests the effect of this. Respondents based locally were found more likely to mention user and existing operator topics, which are more likely to involve local governments and transport operators. In contrast, *Service Planning* topics were likely contributed by state government representatives, while *Future Platforms* were more likely discussed by MaaS or new mobility practitioners. This analysis highlights the influence of respondent characteristics on the topics discussed, offering valuable insights into stakeholder perspectives and priorities related to MaaS implementation in regional Queensland.

Figure 3: Estimated marginal effect of metadata to the six identified topics



4. Discussion

STM proves to be a useful method in quantifying and categorising large amounts of unstructured text into meaningful topical themes. Similar work that was performed manually can be accomplished more efficiently or be used to complement with traditional means of content analysis. This is particularly useful for larger textual datasets. The inclusion of metadata from the responses further aids in identifying the types of respondents more likely to contribute to specific topics and their underlying perspectives. This is quantitatively measured by the marginal effect how each metadata variable impact on the topics identified.

In terms of the study's contribution, the work successfully highlights the various tensions and complexities surrounding MaaS development, such as differing stances on its implementation, the roles of stakeholders (government or non-government), and whether respondents have a local or non-local focus. In the context of MaaS research, several of the topics identified align with those in the qualitative findings, particularly the tension between local and non-local actors. There is a risk that those who are positive about MaaS (tending to be non-local) may impose their system view of transport, overlooking the concerns and needs of local communities. Future MaaS development should strive to strike a balance between local and non-local perspectives, so as to ensure that all stakeholder needs are addressed and integrated in a collaborative manner.

This study serves as an initial step in analysing textual data through STM. Future research could expand on this approach by rearranging responses into other topical themes (e.g., enablers, barriers, or opportunities). Although this could provide an additional dimension of data, some metadata about respondents may be lost during aggregation. Consequently, future work could explore the following directions:

1. Organise the data into different ‘document’ combinations. For instance, the questions asked could be arranged to form a ‘document’ that more effectively overlays respondent groupings and highlights various perspectives.
2. Compare findings with a larger and more diverse cohort of respondents. The research has conducted similar interviews internationally, and these international samples could provide insightful comparisons. STM offers a time-efficient, yet reliable, method for machine learning of human textual inputs across various contexts.
3. Conduct more advanced statistical analyses, such as regression and correlation analyses. This approach could be particularly effective if the study's sample size is increased, providing a larger dataset and adding statistical power to the findings.

5. Limitations and future research opportunities

This exploratory work experimented the use of STM in Australian transport research and demonstrates the potential of STM in various applications. There are several limitations in this work. First, we are unable show the detail individual alignments in the metadata so as to preserve respondent anonymity and maintain research ethics. Further, the number of participants should be enlarged so as to better make use of the capabilities of STM. More work of such can be applied to other stakeholder engagement work, such as the interviews conducted by NSW researchers (Kandanaarachchi et al., 2022). STM can be expanded to analyse any textual data, not only surveys and interviews but also social media data mining and other unstructured sources. Combining datasets from different sources is also possible; as metadata preserves source information, cross-examination of varying sources can be conducted, potentially revealing new insights. If timestamps are available, monitoring temporal trends and changes in user perspectives is also possible.

6. References

Bai, X., Zhang, X., Li, K. X., Zhou, Y., & Yuen, K. F. (2021). Research topics and trends in the maritime transport: A structural topic model. *Transport Policy*, *102*, 11–24. https://doi.org/10.1016/j.tranpol.2020.12.013

Bardutz, E., & Bigazzi, A. (2022). Communicating perceptions of pedestrian comfort and safety: Structural topic modeling of open response survey comments. *Transportation Research Interdisciplinary Perspectives*, *14*, 100600. https://doi.org/10.1016/j.trip.2022.100600

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Das, S., Dixon, K., Sun, X., Dutta, A., & Zupancich, M. (2017). Trends in Transportation Research: Exploring Content Analysis in Topics. *Transportation Research Record: Journal of the Transportation Research Board*, *2614*(1), 27–38. https://doi.org/10.3141/2614-04

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Department of Transport and Main Roads. (2023). *Mobility as a Service—A Queensland Approach*. https://qtenders.epw.qld.gov.au/qtenders/tender/display/tender-details.do?CSRFNONCE=E5F32952C4984BD6648A170534AE26AE

Eckhardt, J. (2019). *Mobility as a Service (MaaS) in rural context*. https://nordicroads.com/mobility-service-maas-rural-context/

Eckhardt, J., Lauhkonen, A., & Aapaoja, A. (2020). Impact assessment of rural PPP MaaS pilots. *European Transport Research Review*, *12*(1), 49. https://doi.org/10.1186/s12544-020-00443-5

Fujisaki, K., Yasuda, T., Ishigami, T., Makimura, K., & Ishida, H. (2022). Empirical recommendations based on case studies in Japan for sustainable innovative mobility in rural areas. *Asian Transport Studies*, *8*, 100079. https://doi.org/10.1016/j.eastsj.2022.100079

Hayauchi, G., & Hidaka, Y. (2022, November 29). *Key Performance Indicators and future perspectives in Japanese MaaS Projects: Current situation in government-sponsored Projects from fiscal 2019 to 2021*. 3rd International Conference on Mobility as a Service, ICoMaaS 2022, Tampere, Finland. https://events.tuni.fi/icomaas2022/pro/

Kandanaarachchi, T., Nelson, J., & Ho, C. (2022). Building Trust and Collaboration Among the Stakeholders in a Mobility as a Service Ecosystem – Insights from Two Maas Case Studies. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4253442

Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies*, *87*, 105–122. https://doi.org/10.1016/j.trc.2017.12.018

Leung, A., & Burke, M. (2021). The prospects for Tourism-focused Mobility as a Service in Queensland. *Australasian Transport Research Forum 2021*.

Leung, A., Burke, M., Akbar, D., & Kaufman, B. (2021). *Mobility as a Service—Regional Research*. Cities Research Institute, Griffith University. https://www.griffith.edu.au/__data/assets/pdf_file/0040/1379947/MaaS_Regional_Research_Report_FINALREPORT_Final1.pdf

Leung, A., Burke, M., & Scott, P. (2023). Tourism MaaS – The case for regional cities. *Research in Transportation Business & Management*, *49*, 101017. https://doi.org/10.1016/j.rtbm.2023.101017

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing Semantic Coherence in Topic Models. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262–272.

Ministry of Land, Infrastructure, Transport and Tourism. (2019). *White Paper on Transport Policy (in Japanese)*. https://www.mlit.go.jp/sogoseisaku/transport/sosei_transport_fr_000098.html

Mladenović, M. N. (2021). Mobility as a Service. In *International Encyclopedia of Transportation* (pp. 12–18). Elsevier. https://doi.org/10.1016/B978-0-08-102671-7.10607-4

Nunez-Mir, G. C., Iannone, B. V., Pijanowski, B. C., Kong, N., & Fei, S. (2016). Automated content analysis: Addressing the big literature challenge in ecology and evolution. *Methods in Ecology and Evolution*, *7*(11), 1262–1272. https://doi.org/10.1111/2041-210X.12602

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association*, *111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: R Package for Structural Topic Models. *Journal of Statistical Software*, *91*(2). https://doi.org/10.18637/jss.v091.i02

Roberts, M. E., Stewart, B. M., Tingley, D., & Airoldi, E. M. (2013, December 5). The structural topic model and applied social science. *Proceedings of the 26th International Conference on Neural Information Processing Systems*. International Conference on Neural Information Processing, Lake Tahoe NV, USA. https://wcfia.harvard.edu/publications/structural-topic-model-and-applied-social-science

Shibuya, K., Kagami, H., Kamiya, H., & Hato, Y. (2022). A Research on the Cooperation between MaaS and Tourism in Japan (In Japanese). *Proceedings of the General Meeting of the Association of Japanese Geographers*, *2022s*, 154. https://doi.org/10.14866/ajg.2022s.0_154

Tamakloe, R., Park, D., & Chang, H. (2022). Discovering research topics, trends, and perspectives in COVID-19-related transportation journal articles. *International Journal of Urban Sciences*, *26*(4), 710–738. https://doi.org/10.1080/12265934.2022.2044891

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. *Proceedings of the 26th Annual International Conference on Machine Learning*, 1105–1112. https://doi.org/10.1145/1553374.1553515

World Economic Forum. (2020). *White Paper—Transforming Rural Mobility in Japan and the World*. World Economic Forum.