# Energy justice, built environment and poverty: A deep-narrative analysis of energy cultures in Brazil, India and Nigeria for policy modelling

Ramit Debnath*[a,b], Ronita Bardhan[a], Sarah Darby[c], Kamiar Mohaddes[b,d], Ana Cristina Villaça Coelho[e], Abdulrasheed Isa[f], Minna Sunikka-Blank[a]

[a]Behaviour and Building Performance Group, The Martin Centre for Architectural and Urban Studies, Department of Architecture, University of Cambridge, Cambridge CB2 1PX, United Kingdom

[b]Energy Policy Research Group, Judge Business School, University of Cambridge, Cambridge CB2 1AG, United Kingdom

[c]Energy Program, Environmental Change Institute, University of Oxford, Oxford OX1 3QY, United Kingdom

[d]Judge Business School, University of Cambridge, CB2 1AG, United Kingdom

[e]Sustainable Buildings Research Centre, University of Wollongong, New South Wales 2522, Australia

[f]Department of Economics, Nile University of Nigeria, Abuja 900001, Nigeria

## Abstract

Distributive energy justice as a policy instrument is crucial to meet UN-SDG 7 targets at the poorest section of rapidly urbanising Global South. This study mobilised this policy instrument to derive poverty alleviation focal points across slum rehabilitation housing in Brazil, India and Nigeria. It emphasises on the social dimensions of energy services which currently remains a literature gap. The core methodological innovation lies in the nested deep narrative analysis approach using topic modelling and grounded theory, developed in Debnath et. al. (2020). This paper presents the application of this methodology through the analysis of energy culture and built environment in poverty. Narratives of poverty and energy services were collected through focus group discussions that served as the primary data object. Topic modelling was performed to remove directionality biases in the narratives using a Latent Dirichlet Allocation (LDA)-based algorithm, which was then strengthen using a grounded theory. Cross-validation of the topic models was performed using science-mapping of state-of-the-art literature in 'energy justice' and 'poverty' with an LDA-tuning algorithm. Results converged around seven key topics that probabilistically highlighted the grounded reality of complex interactions between poverty, built environment and human-scale energy services. The derived policy-focal points illustrated that built environment design, land use planning and governance are crucial to the success of SDG-7 in poverty. It showed that how utilities and distribution companies can strengthen their energy governance to support social policies and improve the quality and reliability of electric power. The high diffusion of ICTs in the study area can act as platform of change to bring distributive justice.

Keywords: energy justice; poverty; computational social science; policy modelling; energy services

## *corresponding author

Behaviour and Building Performance Group, The Martin Centre for Architectural and Urban Studies, Department of Architecture, University of Cambridge, Cambridge CB2 1PX, United Kingdom. Email: rd545@cam.ac.uk

# 1. Introduction

Distributive energy justice as a policy instrument is crucial to meet UN-SDG 7 targets at the poorest section of the society, especially for rapidly urbanising Global South. The forces of urbanisation are mobilising millions of citizens from rural and low-income areas to cities for a better quality of life. However, the cost of this mobilisation is reflected in the ever-increasing slums settlements. Recent estimates show that a quarter of the world's population lives in slums (Hutt, 2016). Energy is regarded as the 'golden thread' that can aid in poverty alleviation by connecting economic growth, social equity and environmental sustainability (Brand-Correa, Martin-Ortega, & Steinberger, 2018). However, energy, which is measured in physical quantities, cannot be weaved to address poverty, but rather it is the benefits that people derive from energy carriers that contributes to well-being (Brand-Correa et al., 2018).

People use energy to derive specific energy services that contribute to their well-being, termed as *human-scale* energy services (Brand-Correa & Steinberger, 2017). For people living in poverty, they are likely to be concerned with the provisioning a range of essential energy services that can improve their well-being (Samarakoon, 2019). Physically these essential energy services must provide the minimum of lighting, cooling, heating, charging and powering appliances, while socio-culturally, it must also satisfy the need for comfort, convenience and cleanliness (Sovacool, 2011). The current energy governance regimes in the Global South are based on top-down institutional narratives focussed on accessing modern energy and arbitrary rates of minimum consumption that do not necessarily always translate into improved well-being effects (Munro & Schiffer, 2019; Samarakoon, 2019). Distributive energy justice provides a conceptual platform to integrate the spatial, temporal and social dimension of energy to maximise welfare benefits from energy services in poverty (Bartiaux, Vandeschrick, Moezzi, & Frogneux, 2018; Bouzarovski & Simcock, 2017; Sovacool & Dworkin, 2014). We believe it can be a useful tool in wielding poverty alleviation policies by appropriating the needed human-scale energy services in the poorest section of society.

In this purview, we perform energy policy modelling in poverty through a distributive energy justice lens in three rapidly urbanising cities of Global South using narratives as the primary data object. Here, the context of poverty in rapid urbanisation is established through the investigation of slum rehabilitation housing in Brazil, India and Nigeria. Slums become the first habitus of the low-income population which migrates to *big cities*. The government often use policy instruments like social housing and slum rehabilitation housing to rehabilitate these people from the slums to formal houses with a vision of improving the quality of life (Bardhan, Sarkar, Jana, & Velaga, 2015; Nijman, 2008). However, market and political-economic forces often impede the trickle-down welfare effects of these social housing (Nijman, 2008; Rondinelli, 1990). Recent studies have shown that poor built environment design of slum rehabilitation/social housing deteriorates the quality of life of occupants, putting them to added poverty burden, including energy poverty (Debnath, Bardhan, & Sunikka-Blank, 2019a, 2019b; Mitlin & Sattherthwaite, 2012; Sunikka-Blank, Bardhan, & Haque, 2019). We investigate this built environment – energy poverty link by exploring energy cultures (after Stephenson, (2018) in poverty in the three cities (Rio de Janeiro, Mumbai and Abuja) of the Global South.

Three narratives of energy cultures in poverty across Brazil, India and Nigeria are then analysed using the deep narrative-analysis methodology developed in Debnath et. al (2020). In doing so, we derive policy focal points that can inform policymakers and local governance in distributive energy justice at the bottom of the pyramid population. The narratives on energy culture in slum rehabilitation built environment in the study area is deconstructed into its word-word, word-topic and topic-topic probabilities through topic modelling, and then interpreted using grounded theory to link the latent social processes, understanding of which is crucial for 'good' energy policymaking (Ozawa, Chaplin, Pollitt, Reiner, & Warde, 2019). The context of text-based narrative analysing of energy culture in poverty using a computational social science perspective for deriving generalised policy focus points defines the novelty of this study. Most topic modelling (TM) research in literature is performed unidisciplinarily-based on big data analytics (Jelodar et al., 2019), this study for the first time attempts to perform a nested TM using a small-data object (text-based narratives) that further adds to the novelty of this study.

The innovation presented in this study is two-folded, first is the nested application of topic modelling and grounded theory in analysing energy culture in poverty. Second is the derivation of distributive energy justice policy focus points in poverty that connects energy services, built environment design and utility governance with poverty alleviation policies (slum rehabilitation/social housing). Besides, this study illustrates the usage of narratives in topic modelling and presents a proof-of-concept of the applicability of quantitative tools (like TM) for qualitative analysis (narrative-based energy policy research). In doing so, the nested model argues for removing directional biases that often arise in narrative analysis, especially in heterogeneous context. Directional biases can reduce effectiveness in the socio-technical policymaking processes (Fukuyama & Weber, 2009). It can simply lead to bad decisions with perverse effects, like uniform energy tariff in slum rehabilitation housing of Mumbai which disregards the socio-cultural complexities associate with poverty (Debnath et al., 2019b). Trotter, (2012) reasons that removing directional biases in qualitative research can improve its policy applications; however, it continues to remain a gap in the current literature. Our nested methodology addresses this gap (discussed in detail in the Part – I) by adopting Bayesian inference in narrative-driven policy modelling.

The primary question that we are trying to answer through the application of the nested methodology is, 'what are the focus points in devising a distributive energy justice policy under the constraint of poverty and rapid urbanisation?'. To answer this question, we employ two boundary conditions. First, we base our energy policymaking dialogue through the lens of energy culture of availing human-scale energy services in the study area. Second, slum rehabilitation housing/social housing are used as a proxy to emulate the living conditions and built environment in urban poverty of the Global South. The specific research objectives of this study are, a) to establish the latent links between built environment design and planning, energy service provisioning and utility governance in low-income communities that can be used to promote distributive energy justice; b) to use narratives as a data object for energy policymaking using the proposed nested deep-narrative analysis framework (see Debnath et. al (2020); c) to evaluate the efficacy of computational social science tools like topic modelling with qualitative analysis for cutting-edge, evidence-based policymaking

This study is divided into the following sections. Section 2 presents a state-of-the-art research background on human scale energy services in poverty and distributive energy justice. Section 3 illustrates the overall methodology of this study concerning narrative collection, data analysis and topic modelling validation using bibliometric mapping. Also, this section also provides an overview of the mathematical background of the topic modelling technique used in this study. Results, discussions and the derivation of policy focus points are presented in section 4. Finally, section 5 offers concluding remarks.

# 2. Background

## 2.1 Energy justice, poverty and energy services

The state-of-the-art research on energy justice and poverty alleviation focus on two broad themes of *essential energy services* and *well-being* (Samarakoon, 2019) that redefines energy as a *"… golden thread that connects economic growth, social equity, and environmental sustainability*" (UN, 2012). In this vein, global efforts to address energy poverty are primarily framed in technical terms that modelled the physical dimensions of 'modern' energy services. Furthermore, achieving minimum rates of energy consumption to improve the health and well-being of people living in poverty (Shyu, 2014; UN, 2018). The search for 'good' heating and cooling policies in the Global North and Global South, respectively, remains an 'energy justice' challenge in the uncertainties of the climate change (Ozawa et al., 2019). Addressing this challenge is more critical in poverty as the well-being outcomes are deeply coupled with the availability of a broad range of socio-economic energy services that happen to require energy for lighting, cooking, cooling, heating, charging and powering appliances. The 'golden thread' embodiment of energy in poverty alleviation, is not an end in and of itself, but rather, a means through which a wide array of 'human-scale' energy services can be provisioned for better well-being (Brand-Correa et al., 2018; Samarakoon, 2019).

Human-scale energy services are referred to the socio-technical dimension of energy needs that extend beyond the physical quantification of energy and power (e.g. in kilowatt-hours, kWh, or kilowatts, kW). The human dimension of energy services relates to socio-cultural drivers of energy service demand, which sociologist Elizabeth Shove categorises as the energy needed for comfort, cleanliness and convenience (Shove, 2003). This concept was further expanded in the theoretical proposition of urban service energy ladder (Sovacool, 2011), and energy culture (Stephenson, 2018). Brand-Correa, Martin-Ortega, & Steinberger, (2018) expands the applicability of such abstract concept of human-scale energy services through the introduction of a community-led participatory approach, called HUSES (Human-Scale Energy Service provision framework). HUSES intended to identify connections between energy services and human satisfaction, such that the community paves its pathways to satisfy its socio-technical needs for energy services. Application of HUSES in local communities of Medellin Columbia has shown that communities derived effective provisioning of energy services of Information and Communication Technologies (ICT) and mobility. It indicates the need for similar human-centric energy service considerations in the current policy regimes as drivers of sustainability and well-being.

Samarakoon, (2019) expanded the understanding between energy services and well-being in the energy-poor of the Global South by forwarding a conceptual framework based on the theoretical impediments of energy justice. The framework is based on two foundational concepts, a need-centred understanding of eudemonic well-being, and viewing energy use through the lens of energy services. The author implied that a strong understanding of the issues of energy justice is needed to provide energy services for poverty alleviation efficiently. However, this study did not fragment energy service into the specificities of social and technical dynamics of energy services that can be referred to an empirical analysis for policy interventions. The empirical representation of the social dimensions of energy services in the Global South remains a critical literature gap (see Fig 1). It is primarily because top-down narratives have long shaped energy narratives in the Global South from institutions in the Global North. This study contributes to this literature gap by computationally analysing energy narratives in urban poverty in three hyper-urban cities of the Global South. In doing so, this study intends to generate a methodology of 'social intelligence' from narratives on distributive energy justice in poverty in a bottom-up manner. It can extend energy interventions that are just and sustainable to the vulnerable groups.

## 2.2 Policy modelling approaches for distributive energy justice

Public policy is often considered as the solution to correct injustices associated with the challenges of energy transition (Sovacool & Dworkin, 2015). However, policies intended to transform the energy system are observed to have justice issues themselves as the energy and climate models have some degree of directional biases in the current regime of energy policymaking by scientists and technologists (Jenkins, Mccauley, Heffron, Stephan, & Rehner, 2016; D McCauley, Heffron, Steohan, & Jenkins, 2013; Ozawa et al., 2019). It is attributed to the justification of technocratic propositions of preferred 'science and technology' complimenting the objectives of the leading political forces (Ozawa et al., 2019). It restricts the scope of climate and energy system models, making them ineffective across the horizontal societal domains. Technology is indeed inevitable to address climate change and sustainable development goals, but policies should also be wielded with the welfare goals that a specific technology can deliver. Thus, resolving the 'last mile' problem in systems planning (Boarnet, Giuliano, Hou, & Shin, 2017; MacKay, 2008). While a growing number of studies have empirically investigated injustices associated with energy and climate policies in the Global North, but they were exploratory and did not converge into the specificities of policy modelling for energy justice. For instance, Grover & Daniels, (2017) spatially explored the distributional injustices associated with the feed-in tariff as a policy instrument for deploying clean energy in England and Wales. They had used cross-sectional regressions to link settlement density, homeownership status, physical dwelling type, local information spillovers, and household social class as crucial non-income variables of distributional energy justice (Grover & Daniels, 2017b).

Similarly, Day, Walker, & Simcock, (2016) have used the capability approach to conceptually scope the policy route to distributive energy justice in energy poverty. It was then expanded in Walker, Simcock, & Day, (2016) in establishing a minimum of energy use in the United Kingdom for addressing energy poverty. The authors have reasoned that establishing a minimum of energy-use is multidimensional that traverses public policy concerning questions of health, social participation, opportunity and practicality (Walker et al., 2016). They argued that public deliberations on bare essentials of energy services should be taken as legitimate grounding for defining minimum standards (Walker et al., 2016). This argument serves as a critical foundation for this study, as we derive policy focus points through public deliberations of living in slum rehabilitation housing and the demand for human-scale energy services. In a recent exploratory study by Zhou & Noonan (2019) on distributive justice implications of clean energy policies in the United States have observed racial and socioeconomic disparities in three government-driven clean energy programs. The authors reported that the likelihood of smart meter rollout is governed by the racial bias of utility companies, and inequalities in funding clean energy programs (Zhou & Noonan, 2019). They also found that improving the government's procurement policy can improve distributive equity for energy-efficient buildings (Zhou & Noonan, 2019). Thus, revealing race and socioeconomic composition as a crucial element of distributive energy justice policy planning.

A recent book edited by Bombaerts, Jenkins, Sanusi, & Guoyu, (2020) titled 'Energy Justice Across Borders' sheds light on energy justice in practice in the Global South. Kruger & McCauley, (2020) conducts a qualitative-based spatial exploration of justice in the Democratic Republic of Congo and the grid systems derived from hydropower. The authors explore injustices in macro and micro typologies to deconstruct the pre-determined Wester logics of justice associated with energy provisioning. Similar conclusions were made by Darren McCauley et al., (2019) at a global scale. We apply similar deconstructivist approach through topic modelling in this study. In the same book, Govindan, Palit, Murali, & Sankar, (2020) argued the need of gender parity and social inclusion in the context of just electricity policies in India, Kenya and Nepal. They assert that the assumption that electricity access itself is enough for associated welfare benefits for both men and women is wrong, and this 'gender-blind' approach in energy policymaking needs to change (Govindan, Palit, Murali, & Sankar, 2020b). They found that the lack of documented evidence on the merit of gender parity in electrification policies and programmes further aggravates gender-blindness in the current regimes of energy policy. Our study contributes to this gap by exploring women voices in energy cultures in poverty.

Tucho, (2020) argues that a significant barrier to distributive energy justice policy modelling is the weak understanding of local contexts and societal needs, and lack of bottom-up innovations. The author calls for gender parity in energy policies for low-income countries such that just energy provisioning materialises in income generation and poverty alleviation (Tucho, 2020b). A conceptual policy modelling approach for addressing energy injustices was presented by Lee & Byrne, (2019), which was then applied to the case of South Korea's nuclear power system and Seoul's One Less Nuclear Power Initiative. The authors argue that in just low-carbon energy policymaking shedding light on underlying drivers of energy injustice embedded in the dominant modern energy paradigm is crucial for energy justice (Lee & Byrne, 2019). However, the studies mentioned above do not forward any methodologies or roadmap for energy justice policy modelling; they strengthen the theoretical foundations of the cross-sectional concepts associated with it.

# 3. Data and methods

The overall methodological framework adopted in this paper is illustrated in Fig 1. The broad methodology consists of three steps, i) application of proposed nested deep-narrative analysis framework from Debnath et. al (2020); ii) topic model validation and verification using science mapping techniques, and iii) derivation of policy focal points based on the deep-narrative analysis.

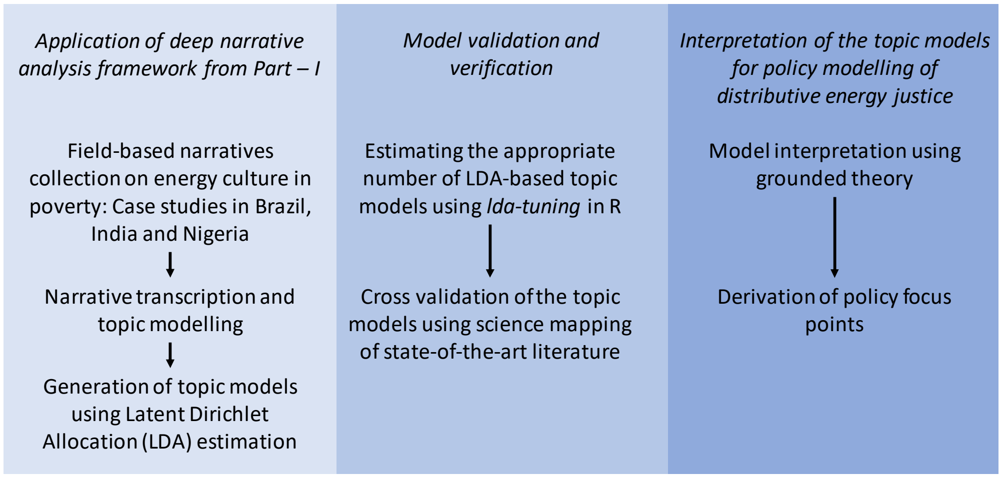



Fig 1. The methodological framework of this paper

## 3.1 Data collection and study area.

Focus group discussion (FGD) was the primary tool of data collection in three distinct slum rehabilitation housing in Brazil, India and Nigeria, in the cities of Rio de Janeiro, Mumbai and Abuja, respectively. FGDs were conducted as per the best practice guidelines of Gill, Stewart, Treasure, & Chadwick, (2008). The interviews were conducted to extract grassroots-level information on human-scale energy service demand in poverty using the lens of energy culture (after Stephenson et al., (2010)). Recently, Tesfamichael, Bastille, & Leach, (2020) have used a similar approach of energy culture investigation in Kenya, and it provides valuable clues for this study to conduct the FGDs. As it offers a coherent explanation as to why electricity consumption in rural households remains low even after connection to an electricity grid, and analyses the value ascribed to different end-uses.

Energy culture theory explores the interaction between energy practices, socio-cultural cultural norms and material reality that shapes the need for certain energy services (Stephenson et al., 2010). Using this epistemology, we encourage informal dialogues with women of the community concerning practices (like appliance use, energy-built environment connections), norms (like the use of a specific appliance) and material reality (like appliance ownership drivers) of energy consumption in the study areas. The open-ended questions of the interviews are illustrated in Table 1. We specifically interview women as the women in such low-income communities spend most of their time in the community, interacting with the built environment, while the male member goes out to work. This gendered approach was influenced by the recent findings of Sunikka-Blank et al., (2019) on gender-built environment interaction in energy demand in slums; and the gap of gender-parity in electrification policies for distributive energy justice, as forwarded by Govindan et al., (2020).

Table 1. Questions to explore energy culture in the study areas through FGDs

| **Energy culture components** | **Open ended questions on** |
|---|---|
| Material reality | • Household appliance ownership<br>• Energy sources and services<br>• Cooling devices<br>• Built environment and architectural characteristics |
| Norms | • Aspirations associated with housing and appliance uptake<br>• Expected comfort levels<br>• Expected changes to built environment and housing characteristics<br>• Gender roles on appliance purchasing decision<br>• Energy bill expectations and realities |
| Energy Practices | • Utility governance in the neighbourhood and tariff mechanisms<br>• Appliance usage and cooking practices<br>• Cooling device usage<br>• ICT device usage and interactions (smartphones and Wi-Fi use)<br>• Energy use and socio-architectural influences<br>• Utility of welfare appliance (refrigerator and washing machine)<br>• Maintenance of devices<br>• Power quality and reliability (load shedding) |

The FGD participants were recruited by us asking a female local contact in each community to bring along some of her friends to a focus group. This method was employed so that a familiar trust was built during the interview process. The total number of participants in Brazil, India and Nigeria were 7, 11 and 6, respectively. The primary occupation of the participants was 'housewife'; however, the socio-cultural construct of the slum rehabilitation areas usually encourages the women to engage in informal income generation activities. For example, in Mumbai, many participants engage in tailoring and garland-making to support their livelihood. The education level varied in the three cases, in Brazil most of the participants were primary school dropouts. In India, the education level of the participants was middle school dropouts. And, in the Nigerian case none of the participants attended formal schooling.

All members of each FG spoke the same mother tongue that aided in a good conversation. The moderators in the three FGDs were bilingual (mother tongue – English) who transcribed the interviews from the local language to English. In Mumbai, India the translation was done from *Hindi – English*; in Rio de Janeiro, Brazil, *Portuguese – English* and in Abuja, Nigeria, *Hausa – English*. In qualitative coding, the doubt arises with such heterogeneous transcription is that how far we can transfer concepts and terms across languages and feel confident that they mean the same thing. The multilingual nature of the process will involve uncertainties about meaning that may induce directional biases in its interpretation (Bryant & Charmaz, 2019; Trotter, 2012).

The moderators of the FGDs were part of the community. They were community outreach workers who have had previous experience in conducting FGDs. It helped in setting up the atmosphere of the interview as the participants were familiar with each moderator in the three case areas. Each moderator briefed the participants over the communal tea-time on the objective of the FGD and our quest for a better understanding of energy demand in poverty that aided in establishing a suitable vocabulary for the FG questions. A picture from the FGD is illustrated in Fig 2. The moderator obtained verbal consent from each of the participants before the start of FGDs. The socioeconomic, energy use and built environment characteristics of the study areas are shown in Table 2 that shows the contrast in the study areas within the similarities of slum rehabilitation. For example, Indian case is a vertical slum rehabilitation that restricted access to public/open spaces, in contrast to the Brazilian and Nigerian cases. The Indian households had access to metered electricity grid connections. The Nigerian household had grid connections, but meters were non-uniform across households. Contrastingly, the Brazilian households did not have any formal grid connections or meters, occupants directly used electricity from overhead electric transmission lines through metal hooks. Such heterogeneities in the slum rehabilitation built environment illustrates the spatial characteristics of distributive energy injustices.

Fig 2. A focus group discussion in progress in Mumbai, India

Table 2. Socio-economic, built environment and energy-use characteristics of the study area

| **Characteristics** | **Brazil** | **India** | **Nigeria** |
|---|---|---|---|
| **Built environment** | | | |
| City | Rio de Janeiro | Mumbai | Abuja |
| Percentage of population living in slums (as per 2019 data) | ~ 24% | ~ 41% | ~ 79% |
| Slum rehabilitation housing typology | Horizontal, single storey | Vertical, multi-storey | Horizontal, courtyard-type |
| Slum rehabilitation case | Displaced due to loss of employment in waste-recycling industries | Displaced from horizontal slums due to political-economic forces of urbanisation | Rural to urban migration, forced resettlement |
| Number of rooms | 2 | 1 | 1-3 (variable) |
| Floor area ($m^2$) | ~20 | ~25 | ~24 |
| **Socio-economic** | | | |
| Average household income (monthly) | 80 – 160 USD (when employed in the waste-recycling industry) | 70 – 140 USD | 70 – 100 USD |
| Primary earner | Head of household (was the husband) | Head of household (The husband) | Head of household (The husband) |
| Present primary occupation of the HoH | Unemployed at the time of survey and dependent on charity | Daily wage labourer in construction industry, security guard, watchman, taxi-driver | Daily wage labourer, taxi-driver, security guard |
| Average number of people per household | ~5 | ~5 | ~7 |
| Informal economic activities | The female members (adults) work as domestic help. The job is irregular. | Few households offered tailoring and garland-making services, run by the wives. | Households with refrigerators, offered water and cold drink cooling services for a fee. |
| **Energy-use** | | | |
| Grid connected | No, but distribution transformer is present | Yes, some buildings have community owned rooftop solar PV for water pump operations | Yes |
| Highest household electricity bill (2018-19) | NA | 112 USD | 83 USD |
| Average household electricity bill (monthly) | No billing (No grid connection). Electricity stolen from overhead transmission lines (220V lines) | 7 – 10 USD (Grid connected). Every household is metered. | 8 – 15 USD (Grid connected). Meters are non-uniform across the houses. |
| Special electricity tariff program | NA | NA | NA |
| Cooking fuel | LPG, electricity | LPG, electricity, kerosene | Firewood, cow dung, biomass-based |
| Typical appliances in the surveyed households | Fans, light bulbs, television (TV), oven, smart phones, refrigerators | Fans, light bulbs, television (TV), smart phones, Wi-Fi routers, refrigerators, washing machine | Fan, light bulb, mobile phones, power adapters |
| **Photos of the built environment** |  |  |  |

### 3.1.1 Nested deep-narrative analysis

The nested deep-narrative analysis framework proposed in Part I of this study (see Fig 1, Debnath et al (2020)) is applied in this paper using a text-based narrative dataset from slum rehabilitation housing in three cities of Global South. The narratives are collected through focus group discussions (FGDs). We argue that the nested approach (see Fig 3) responds to eliminate the directional biases by performing topic modelling at the pre-narrative analysis stage using the transcribed text as the primary data object. In doing so, as explained in Debnath et al (2020), the text data corpus is broken into its fundamental element, i.e., words. These words are then arranged or clustered via LDA according to the hyperparameters of the topic models in a probabilistic manner. LDA represents the clustering as word-word, word-topic, and topic-topic probabilities. This is the application of AI in text mining, and we do not stop our analysis here; instead, we interpret the topic models as per grounded theory to reveal the latent processes in the energy culture of the study areas. We believe, to the best of our knowledge, such an approach in the narrative analysis was never made before, and it demonstrates the methodological of this study.

The applied deep-narrative analysis framework is shown in Fig 3, where the topic modelling was performed using Latent Dirichlet Allocation (LDA) estimation, using a model that allows sets of words in a transcript to be attributed to particular topics. A full rationale for choosing this mode of analysis is presented in Section 2.2 of Debnath et al (2020)).

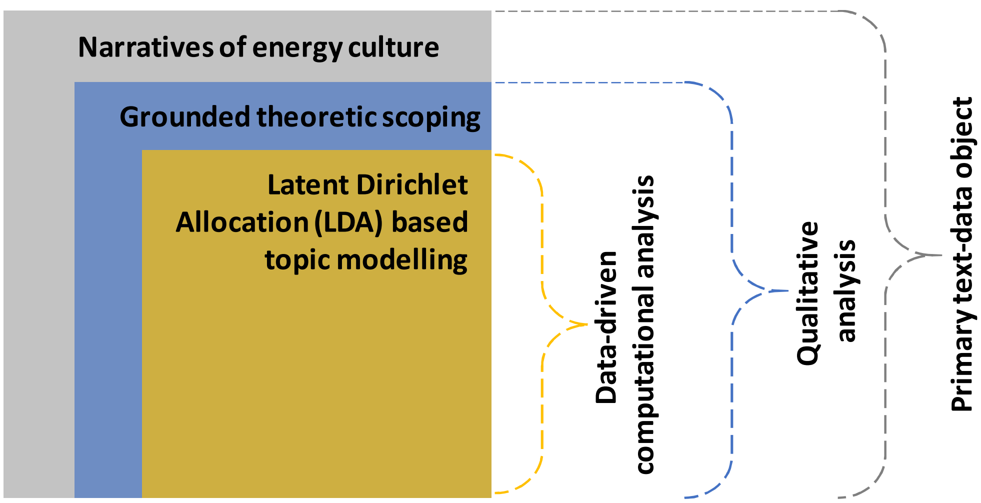



Fig 3. Applied nested deep-narrative analysis framework from Debnath et al (2020)

## 3.2 Topic modelling using Latent Dirichlet Allocation (LDA)

The LDA computation was done by a publicly accessible toolkit called *tidytext*; computational tools are written in the R computer programming language (Silge & Robinson, 2019). The steps followed in our topic modelling and text analysis is as follows: (1) pre-processing, (2) determining the number of topics, (3) setting the control parameters, (4) validation (5) visualisation. The process flow of the topic modelling approach is illustrated in Fig 5. The pre-processing stage consisted of the construction of a 'Document-Term Matrix (DTM)' that structured the text data into a *corpus* as per the specifications of the *tidytext* R-package, called *tidy data* (see Chapter 1, (Silge & Robinson, 2017)). Each sentence was treated as a document in the DTM, that resulted in 504 sets of unique documents (***M***) that had ***w*** (words) and ***z*** (topics) as per LDA model specification (see Appendix A2, Fig A3). Stop words (e.g., 'the', 'of', 'from', words that carry limited information) and punctuations were removed from the corpus using a process called *lemmatization* (Manning, Raghavan, & Schutze, 2009). Lemmatization also involved removal of inflectional ending of words, and converting the grammatical form of a word into the base or dictionary form (known as *Lemma*) (Manning et al., 2009). The corpus object now contains only the base or dictionary form of words arranged in the form of *tidytext* DTM for topic modelling using LDA. Parameter estimation was performed using variational expectation-maximization (VEM) algorithm in the *tidytext* package (Robinson & Silge, 2019). Further processing of the dataset and visualisation was performed using *dplyr* (Wickham, François, Henry, & Müller, 2019), and *ggplot2* (Wickham, 2016), respectively.

Three different comparison metrics were used for determining the number of topics using an R-package, *lda-tuning* (Moor, 2019). A similar LDA-model validation approach was used by Li, Rapkin, Atkinson, Schofield, & Bochner, (2019). The topic numbers were also indicated by the exploratory keyword co-occurrence analysis through science mapping (see Fig 1). In addition, the perplexity measures (Hankerson, Harris, & Peter D Johnson, 2003) of the topics were also tested for cross-validation using the built-in perplexity tool of *tidytext* package.

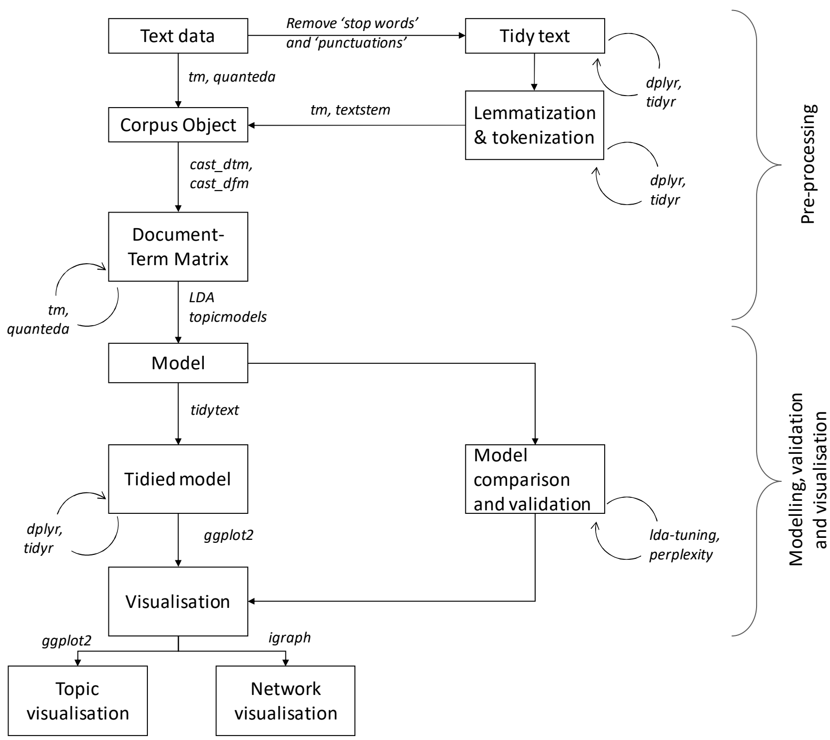



Fig 4. Process flow of LDA-VEM based topic modelling in R [Note: the text in italics represent the R-library in use]

## 3.3 Science mapping for validating topic models

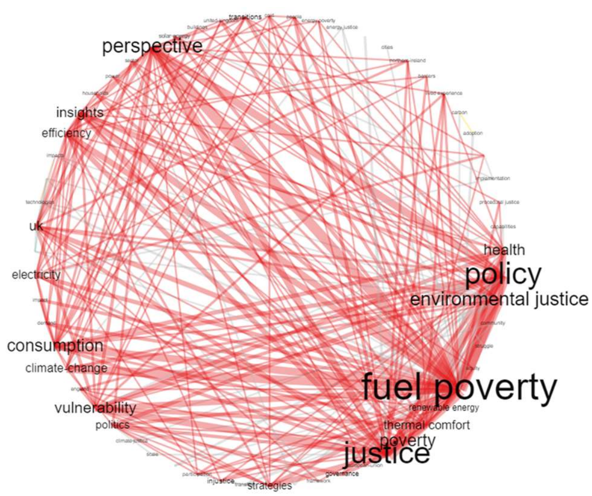
Identification of relevant topics is crucial for the interpretation of topic modelling. We identify it by a cross-validation procedure using high frequency topic in current literature (see Fig 1). The first step is to run *LDA-tuning* package in R and identify convergence points (after the methodological steps of Li et al., (2019)). And, the second step was to cross-validate using high-frequency topics from bibliographic databases on 'energy justice' and 'poverty'. In this study, we have used the Web of Science (WoS) database to identify co-occurring and high-frequency keywords in the state-of-the-art literature. This procedure of identifying co-occurring keywords from bibliographic databases is called science mapping (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011).

The WoS search showed 144 published documents on 'energy justice' and 'poverty' between 1997 – 2019. The keyword selection was kept as broad as possible to sample maximum relevant documents. The keywords were extracted from the 'Title', 'Abstract', 'Author keywords' and 'Journal keywords (Keyword Plus)' of the 144 published papers. The co-occurrence analysis of the bibliographic database was performed using the *bibliometrix* package in R (Aria & Cuccurullo, 2017). Keyword co-occurrence analysis is generally used to provide a graphical visualization of potential relationships between keywords from the titles, abstracts and citation contexts. The *bibliometrix* package uses a network analysis approach to map the keywords based on their co-occurrences in the database. Detailed methodology on bibliometric network analysis using the *bibliometrix* package can be referred from Aria & Cuccurullo, (2019).

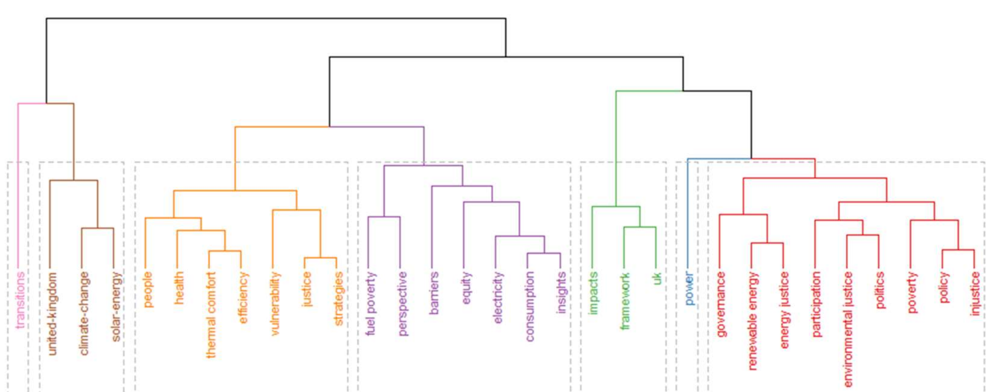
# 4. Results and discussion

## 4.1 Topic co-occurrences in bibliometric database

This section presents the exploratory results of the bibliometric keyword co-occurrence analysis of Web of Science (WoS) database of published articles between 1997 – 2019 (i.e., n = 144). Results at 10% significance-level show the possibility of 7 to 8 key topics in the state-of-the-art literature (see Fig 5). The major topics are identified through the highest frequency keywords in Fig 6a as 'policy'; 'environmental justice'; 'fuel poverty'; 'poverty'; 'justice'; 'consumption'; 'climate change'; 'perspective'; 'health'. However, other repeating keywords are formed around seven clusters, as shown in the topic dendrogram in Fig 6b. This exploratory bibliometric network analysis (also known as science mapping) is critical in deciding the appropriate number of topics topic models (see section 4.2). Further information on the thematic mapping and topical growth pattern in the WoS database is presented in the appendix section (see A1).




a) Keyword co-occurrence network of published documents (n = 144)
[Note: High frequency topics are denoted by the weight of the topics]




b) A topic dendrogram representing seven distinct clusters around high frequency keywords

Fig 5. Keyword co-occurrence network illustrating high frequency keywords on the topics 'energy justice' and 'poverty' in the Web of Science database published between 1997-2019 (n = 144)

## 4.2 Model verification and validation

A 505 x 1317 document term matrix (DTM) was generated with 99% sparsity, following which a series of LDA models were fitted to validate the desired number of topics on energy justice and poverty links. An exploratory bibliometric-based investigation of high occurring keywords in the published database of Web of Science (WoS) revealed around seven high-frequency topics (see Fig 5 and section 4.1). The Griffiths & Steyvers, (2004) metric indicated a suitable model with 6 to 8 topics, as shown in the curve plateau in Fig 5. The Cao et al., (2009) metric indicated approximately 5 to 6 topics with global minima. The metric by Arun, Suresh, Veni Madhavan, & Murthy Narasimha, (2010) showed a monotonically decreasing pattern with a relative plateau around 5 to 7 topics (see Fig 6). A cumulative consideration of all the metric and the exploratory bibliometric analysis, a 7-topic LDA model was used in the subsequent analysis.

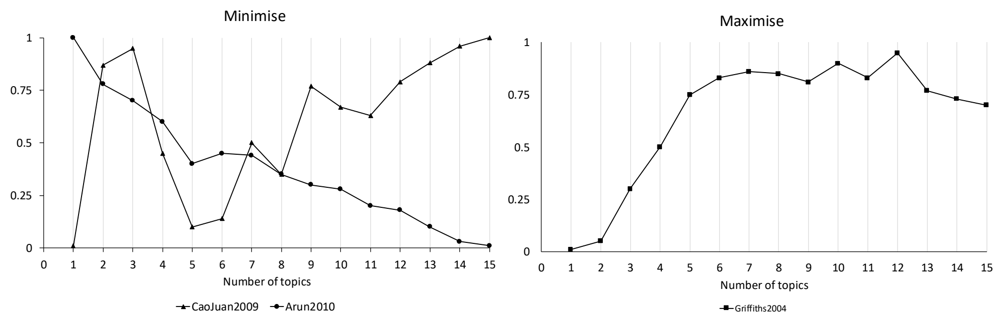



Fig 6. Model comparison metrics to estimate the desired number of topics using a combined text-corpus of narratives from Brazil, India and Nigeria.

[Note: "Griffiths2004" refers to Griffiths & Steyvers, (2004) metric; "CaoJuan2009" refers to Cao et al., (2009) metric and "Arun2010" refers to the metric by Arun, Suresh, Veni Madhavan, & Murthy Narasimha, (2010), as per the taxonomy of *lda-tuning* R-package]

## 4.3 Model interpretation and deep narrative analysis

### 4.3.1 Characteristics of energy culture in the study areas

The energy culture in poverty characteristics in the three cities under study has heterogeneity concerning energy practices, norms and material culture that relates deeply with the typology of the built environment. However, distinct common words on human scale energy services across Brazil, India and Nigeria is represented in Fig 7, that illustrates the most frequent word in the combined narratives on energy culture in poverty. The common thread that connects each case study and their energy culture are the non-income drivers of appliance ownership (as discussed in (Debnath et al., 2019b)). Norms and aspirations concerning appliance ownership are one of the salient features of energy culture in slum rehabilitation housing, as the slum dwellers get a permanent home, their aspirations for a better quality of life increases. It is implicated through higher appliance ownership (see Fig 7), which in turn change the energy practices in the household. This observation is reported by Debnath et al., (2019b) for slum rehabilitation housing in India.

Focussed group discussions revealed that across the three case studies, the energy practise associated with an appliance use is distinct and is strongly influenced by the socio-economic conditions of the place. It, in turn, influences the energy-use norms that dictate the demand for specific energy services. These salient features of energy culture in the study areas are illustrated in Table 3.

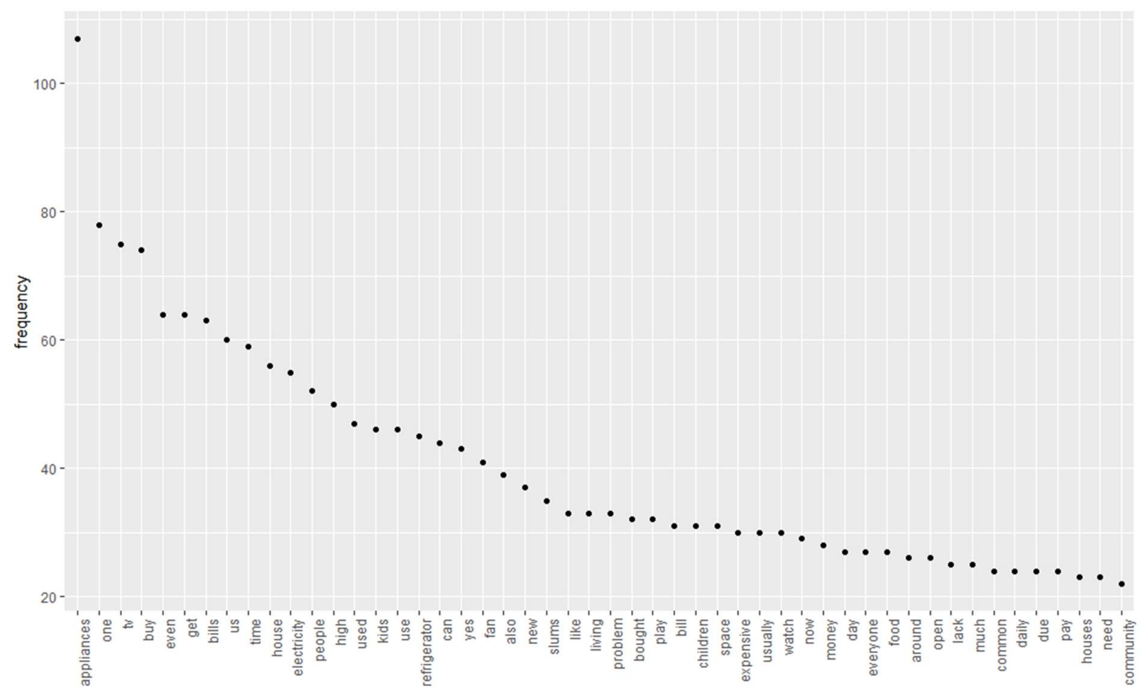



Fig 7. High frequency words from the energy culture narratives in poverty of Brazil, India and Nigeria

Table 3. Descriptive of energy culture in slum rehabilitation built environment in the study areas

| **Energy culture** | **Rio de Janeiro, Brazil** | **Mumbai. India** | **Abuja, Nigeria** |
|---|---|---|---|
| Practices | - Mixed-use of indoor and outdoor spaces for cooking and washing clothes.<br>- TV watching during the day is a common practice.<br>- Energy service demand for baking, cooling using multiple fans, charging phones, lighting, entertainment (TV) and refrigerator use.<br>- Outdoor open spaces are utilised for thermal comfort when indoor temperature becomes very hot. | - Cooking and washing are all performed indoors due to lack of open spaces.<br>- Common mode of watching TV series is on smartphones.<br>- Energy service demand for cooling through fans, lighting, smartphone charging, entertainment (TV and Wi-Fi), washing machines and using refrigerators.<br>- Lack of outdoor open spaces results in higher thermal discomfort during hot summer days.<br>- Unhygienic neighbourhood conditions (like garbage disposed in the open spaces) pushes activities indoors, increasing the energyuse | - Mixed-use of spaces for cooking and washing clothes.<br>- Limited TV-sets, families watch TV communally.<br>- Energy service demand for lighting, cooling through a single fan and charging mobile phones. Only few houses had a TV and refrigerator.<br>- Outdoor open spaces are often used for community activities (like watching TV series) and maintaining thermal comfort. |
| Norms and aspirations | - Aspiration of air conditioner (AC) ownership is high due to hot climate. There is also a high aspiration of smartphone ownership.<br>- Lack of employment forces the occupants to depend on charity.<br>- Most appliances are pre-owned, and refrigerators are received as donations.<br>- Lack of toilets and proper sanitation is a health vulnerability. | - High aspiration of modern appliance ownership for better comfort, convenience and cleanliness, representing middle-income consumption pattern.<br>- Multiple utility of appliance ownership for increasing informal income generation. For example, refrigerators are also used to store fresh flowers that is later sold to garland-making market.<br>- High utility from ICT devices like smartphones and Wi-Fi service for the women of the house. It is used for online shopping and watching TV series. Here, ICT devices have a strong gendered advocacy. | - Aspiration of owning just the bare minimum for subsistence.<br>- Occupants have a strong rural-living characteristic that organises community-sharing of appliances.<br>- TV is watched as a social activity in the community. Refrigerators are owned by only few families that levies a fixed charge for drinking water cooling services in summer months.<br>- People tend to derive higher utility from refrigerators, they store cold drink, and sell it the community during hot days. People must pay to store food and drinks in the refrigerators that are owned by only few households. |
| Material reality | - No electricity meters in the houses, occupants steal it from nearby electric poles.<br>- Appliances are either handed down from higher income households or are received as a part of donation drive.<br>- Lack of toilets, sanitation and hygiene are the most pressing concerns. It has a heavy health burden on the children.<br>- Drug abuse is also a major problem, and the poor quality of built environment enables the proliferation of the drug abuse. | - Utility companies have irrational tariff system that pushes occupants to energy poverty. There is no grievance redressal system with the utility administration and is a major cause of distress in the neighbourhood.<br>- Lack of social spaces in the built environment is a major well-being concern. Occupants say that it is makes them lonely and they must counter it through higher TV-watching.<br>- Lack of safety and hygiene is a serious concern in the built environment. Like illegal parking in the open spaces invites unwanted allied activities and people which makes the neighbourhood unsafe. the spaces between the buildings have become garbage dumping grounds for people living in the higher floors where sanitation services like garbage collection are often skipped.<br>- High diffusion of ICT devices and occupants avail services from it. | - Poor power quality and reliability is a significant issue. Heavy use of power stabilizers that damages appliances.<br>- Lack of responsiveness from utility companies concerning high and irrational billing and frequent load-shedding.<br>- Strong community living that enables sharing of specific appliances like TV and refrigerator.<br>- Lack of enough indoor living space is a major built environmental concern. |

### 4.3.2 Interpretation of topic models and deep-narrative analysis

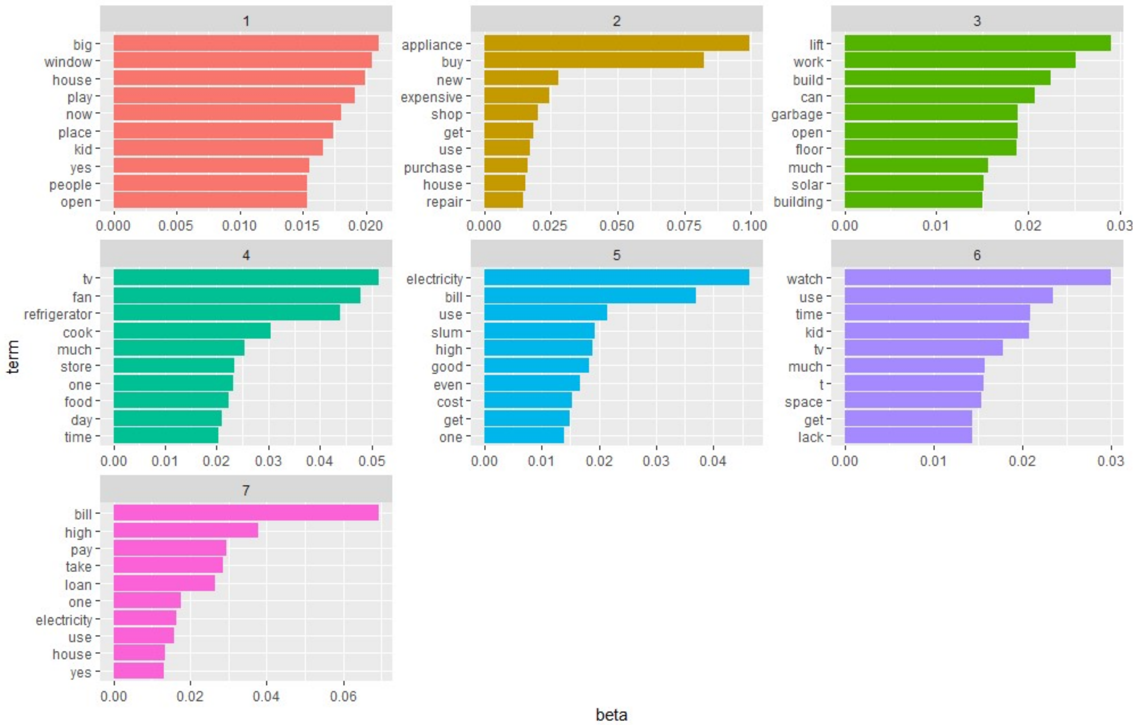



Fig 8 Topic extracted by LDA representing high probable topics concerning energy culture, poverty and energy justice in Brazil, India and Nigeria. [Note: 'beta' represents per-topic-per-word probability]

Fig 8 shows even topic models that best fit the verification criteria, as illustrated in section 4.2. The highlighted words are the top 10 high probability words in each of the topics. These topics are now expanded through a grounded theoretic lens that connects built environment, human-scale energy services and energy culture (see Table 3) in the study areas. The topic 1 and topic 6 appear to be related to the socio-spatial dynamics of the slum rehabilitation built environment (see Fig 8). It further accentuates the built environmental needs for low-income houses with efficient windows for ventilation, open and social spaces for a better quality of life. The word-word distribution in topic 6 also shows a probable relationship between high TV viewing (a critical energy practice) by children in the absence of public spaces in such low-income built environments (discussed in Table 3). Evidence from narratives in Mumbai, India shows this poor built environment – energy demand on children's well-being link:

> *"… the kids are the biggest sufferers here; they do not have any place to play. They cannot play on the staircases because they can fall hurt themselves… due to lack of playground here, my son ran out of the house…. He is just 5 years old, and he said he misses his friends and open areas of slums... A taxi driver caught him crying all alone near our previous home, and then bought him to us…"* (R10, Mumbai, India)

> *"… we cannot share our daily life stories with anyone that's the biggest cause of stress… No spaces for us to sit and chat, neither children have any space to play…they just stick to the TV or mobile phones, either watching cartoon or playing games. It is making them lazy and dumb…"* (R2, Mumbai, India)

> *"… well-being is affected for sure, in addition, the community support system has gone weaker. Previously in the slums, if anything bad happens to any member of the community the information used to travel very fast… everyone used to contribute in supporting the victim. Now, no-one knows their neighbour…"* (R11, Mumbai, India)

In the Brazilian case, occupants were displaced due to closure of a waste recycling site that was their primary job place. Their rehabilitation houses are not provided by the state (unlike the slum rehabilitation houses in Mumbai, India and Abuja, Nigeria). It is impacted through poor sanitation and lack of a toilet in the built environment that has extended health and well-being effects,

> *"… I cannot have guests…we do not have toilets…condition is same as previous"* (R3, Rio de Janeiro, Brazil)

> *"… I cannot even have family guests because I don't have an appropriate place for them to relieve themselves… diarrhoea is common here….my daughter gets frequently ill as I don't have a bathroom; I relieve myself in the bush. I'm telling the truth…."* (R2, Rio de Janeiro, Brazil)

> *"… we have to spend all day outside because the heat inside the house is not good…fans are useless"* (R4, Rio de Janeiro, Brazil)

In the Nigerian case, the rehabilitation is from rural to urban areas, and reflect different social dynamics with the built environment. The primary dissatisfaction with the houses is they are small, and the houses are not allocated based on the family size (as revealed by R2). More interestingly, a significant portion of the room is used for the storage of energy (firewood) for cooking, even though they have access to other forms of cooking energy. It reveals an energy culture of energy stacking in the urban built environment, thus, highlighting a distinctly human-scale energy service – built environment design dynamics,

> *".. rooms are small, they vary in number… it is very difficult to get the room you want as the process of allocation is done by the government. The rooms are very small for a family of 3 and usually most of us are a family of 5…"* (R2, Abuja, Nigeria)

> *"… even in the small room we must keep space for firewood for cooking during the monsoon and winter, or else we do not get cooking fuel. The rooms are not the problem, but their small size is…"* (R1, Abuja, Nigeria)

Topic 1 also showed the significance of 'open' spaces and practical design (need of bigger windows for ventilation) in regulating occupants' well-being and thermal comfort (see Table 3) in low-income built environment (see Fig 8). It accentuates the greater need for built environment design consideration in 'good' energy policy design. Further exploration of topic 6 (see Fig 8), shows high probability words as 'watch' and 'TV', it has a broader significance on the role of technology diffusion of TV and mobile phones that have increased the screen-time. Occupants said lack of space had made them lonelier that forces them to use smartphones for daily entertainment (see Table 3). It had created a dependence on Wi-Fi and ICT-technologies, indicating higher technology diffusion of ICT in such low-income communities,

> *"… I do not get much time to watch TV… My son watches it continuously with my husband. I do not get to watch program of my choice, so I watch it on smartphone through Wi-Fi…"* (R3, Mumbai, India)

> *"…I have a smartphone with Wi-Fi, but I only use it to know the news…"* (R6, Rio de Janeiro, Brazil)

> *"… mostly men and children watch the TV… I often use the phone to watch YouTube or listen to music…"* (R4, Abuja, Nigeria)

This evidence further shows the pace of current technology diffusion of ICT devices in the poorest section of the society. Availing internet services through Wi-Fi informs on the changing energy culture and the kind of energy services that are currently derived. A participant in the Indian FGD even added the convenience of online shopping and the money she saves from added discounts and transportation costs (see Table 3). Therefore, ICT devices can have considerable welfare effects in poverty and should be utilised as a policy tool for distributive energy justice.

Topic 2, topic 4 and topic 5 reveal information on the energy culture and appliance uptake in the low-income built environments under study (see Fig 8). Change of household practices due to slum rehabilitation was found to be associated with higher energy intensity through higher appliance uptake in Mumbai (Debnath et al., 2019b). The probabilistic distribution in topic 2, 4 and 5, respectively, further realises this finding. The energy culture associated with appliance uptake in slum rehabilitation is highly intertwined with the energy services that are derived in poverty (see Table 3). For example, in the Indian case, appliances like refrigerators were bought to derive greater economic utility from improving household welfare. It ranges from storing goods like flowers (see R3 below) to levying charges for availing cooling services (see R5 below). It indicates that welfare appliance like refrigerators has a more significant role as subsistence devices than 'cooling' only device in poverty. A distributive energy justice would mean enabling higher refrigerator ownership in poverty through better policy support.

> *"…food is prepared three times a day and no one stores it in the fridge… common items that are stored are vegetables, milk and water…item like flowers and garlands are also kept as they stay fresh longer, and are then sold in the market..."* (R3, Mumbai, India)

> *"…in the slum, out of 450 houses, only one corner shop had a refrigerator… the shop owner used to charge for using that fridge… Owing a fridge has made my life easier by saving a lot of time…"* (R5, Mumbai, India)

Similar but non-economic utility derivation from refrigerator was observed in the Brazilian case (see R8 below) that indicates multimodal utility of appliances in low-income communities.

> *"…my son likes to keep everything inside the fridge: sugar, rice, coffee. I tell him there is no need for this, but he insists…"* (R8, Rio de Janeiro, Brazil)

While in the Indian case, refrigerators were bought first-hand as an 'aspirational' buy, in the Brazilian case refrigerators were received in the form of donations. On the contrary, the ownership of the refrigerator is highly communal in the Nigerian case. The participants stated that they have an active community that share everyday appliances. However, refrigerator ownership remains scarce as the power quality is poor, and occupants suffer from frequent load shedding (see R4 and R5 below).

Interestingly, they also revealed that instead of the traditional fridge, people in the community opt for deep freezers so that they can store cold drinks and ice-cream (see R below). It highlighted an energy cultural norm of sharing appliances and social uptake of 'utility-specific' appliances like the freezers that can generate an income.

> *"… I have a fridge, and usually share it with the neighbours. Load shedding is a major problem which do not allow the fridge to run properly, if it breaks, I cannot afford another one… We share appliances as required, we have a strong community…"* (R4, Abuja, Nigeria)

> *"… people often buy deep freezers to sell cold items and drinks…they make a good business…"* (R1, Abuja, Nigeria)

> *"…not everyone allows sharing fridges as electricity cut is very common, they need to store food for longer time in hot days…. the electricity bills are too high, and no one wants to let other store it for free…"* (R5, Abuja, Nigeria)

Thus, the energy culture of deriving economic utility from a refrigerator in Indian and the Nigerian cases were viewed as a return on investment by the users. Whereas, in the Brazilian case, the economic utility of the refrigerator was absent may be due to poor-quality of refrigerators received as donations. It is a salient characteristic of Brazilian e-waste recycling industry that every electronic product follows a chain, passing from hand to hand before becoming a disposable waste. It had shaped the energy culture of using pre-owned appliances in the region through donations. It is typical for used equipment and electronic appliances to be given to poorer people in the community (Scartezini, 2013). It further raises the question of energy efficiency of pre-owned appliances. Distributive justice perspective would indicate better energy efficiency in poverty so that the users do not have to pay higher operating costs of using pre-owned appliances.

Fans and TV remain high probability words (see Fig 8) throughout topic 2, 4 and 5 as they have become the minimum requirement of urban living in the Global South. The word 'bills' have a higher probability in topic 5, and topic 7 (see Fig 8), as high electricity bills were the most common complaint in the Indian and the Nigerian case (also mentioned in Table 3). In India, the occupants revealed that frequent shuffling of electricity distribution company causes administrative lags that result in high and irregular bills (see R3 and R4 below).

> *"… there is always huge variation in the bills, sometimes we get Rs. 1300 (~19 USD) and sometimes Rs. 2000 (~28 USD). But it is usually on the higher side and extremely irregular…I got my latest bill after three months…"* (R3, Mumbai, India)

> *"… many people complained to the distribution company about high bills, my neighbour got a bill of RS. 10000 (~140 USD) in a month… no action was taken by them to reduce the bills. The company told us to pay the exact bill, and it will be later resolved in the successive months…but no one here can afford to pay it…"* (R4, Mumbai, India)

The range of bills usually accounts for 40 – 50 % of their monthly household income, pushing them into energy poverty. Similarly, for the Nigerian case, frequent power cuts and poor power quality damages appliances, and it causes significant economic distress in the household (see R1 and R2, below). Occupants spend more money to replace the damaged appliance by buying a brand-new product that creates a poverty trap (further detail on poverty trap can be found (Aklin, Bayer, Harish, & Urpelainen, 2018)). Participants revealed that even after frequent power-cuts they get very high bills, almost 90 - 150% of their monthly households' income (see R3 and R5, below). Surprisingly, they are yet to have electricity meters.

> "*… the bill is too high…sometimes we can go for up to a week without light but at the end they would still bring a bill of around NGN 20,000 (~55 USD) or NGN 30,000 (~83 USD)…it is more surprising that with just one fan and two bulb the bills can be so high…"* (R1, Abuja, Nigeria)

> *"…there was a month one of us was given a bill of NGN 50,000 (~139 USD) by the distribution company…we went to their office to complain and they said we should write a letter and they would address the problem. They are yet to resolve the issue…*
> *… we don't have meters. They are just bringing bills anyhow without measuring our electricity consumption. For example, one of my houses has collapsed long time ago but they keep sending us bills even yesterday although no one lives in the house. That bill is now up to over NGN 200,000 (~554 USD) …"* (R5, Abuja, Nigeria)

> "…*I get very high bill of around NGN 90,000 (~250 USD) due to TV usage. This is very high, and we have complained many times to the distribution office, but nothing is solved…"* (R3, Abuja, Nigeria)

> *"…though the company has cut off the electricity, they would still provide a bill even though we have no electricity… last month the coil of my fan was destroyed due to load shedding, which remains unrepaired... It is too much effort and cost to repair an old fan …"* (R2, Abuja, Nigeria)

The Brazilian case did not have such a problem because they did not have formal electricity connection. The occupants steal electricity from the overhead electric poles. From a distributive energy justice planning perspective, the narratives from India and Nigeria show the governance flaws of the utilities that have severe ramifications in poverty. Poor governance by the utility companies represent injustices in the current energy system that is often ignored in the macro-policymaking. Improving power quality and reliability is as important as the provisioning of power, and it is a critical policy focus point in low-income communities. It is crucial for universal access to energy as per UN-SDG – 7.

High probability of words like 'loan', 'bills', 'electricity' and 'house' in topic 7 (see Fig 8) further emphasise on the poverty penalty of living in slum rehabilitation housing due to injustices caused by poor governance and tariff planning of the utilities. Additionally, in all the three case studies, occupants revealed they had to travel far to repair or purchase appliances that add to the poverty penalty of living in such a built environment. Moreover, there was no trusted pre-owned appliance shop in the Indian and Nigerian case and no culture of handling-down of used appliances (like in Brazilian case). It further illustrates that occupants had to pay the full price along with travel and daily wage costs to purchase an appliance. It shows the nexus of poverty penalty and human-scale energy services that people living in urban poverty experiences. Just energy policy should address this nexus to reduce poverty penalty.

> *"…. used items in general have inferior quality and they need to be repaired frequently. It is an economic burden and a hassle, and it will probably cost us the same of purchasing a new appliance…."* (R11, Mumbai, India)

> *"…children want latest model of the devices, so it's better to buy new products…"* (R3, Mumbai, India)

> *"… It is not only our aspiration; it has become our need. Earlier in slums, even if we aspire to buy the temporary conditions of the house did not permit to purchase such expensive items….*
> *… we need to travel more than 5 kilometres to purchase appliances…."* (R2, Mumbai, India)

> *"… the nearest store is in the centre of Duque de Caxas… which is almost 10 km from here…"* (R1, Rio de Janiero, Brazil)

> *"… there was a market in nearby Kubwa village where we used to buy items. For big purchase like appliances we used to go to town…"* (R1, Abuja, Nigeria)

> *"… we must go far for big shops. The community shops here provide basic material…"* (R5, Abuja, Nigeria)

Until now, the topics (see Fig 8) revealed latent energy cultural attributes that are intrinsically linked with the poverty penalty of living in the slum rehabilitation housing. We also examined the community-grown solutions in the study area. The Brazilian and Nigerian case did not have any such program, but the Indian case illustrated that solar energy led community initiatives to improve the energy service landscape. The topic 3 (see Fig 8) illustrates the solar led 'lift' renovation project in the community to improve the mobility of people in the high-rise neighbourhood (see Table 2). The initiative was initiated by the women-led co-operative in the community who feel their mobility is reduced due to high-rise and no-lift situation in their slum rehabilitation housing (see R3 below). It disrupted their social network. The joint community effort of powering public services through solar energy illustrates distributive justice actions at the grassroots-level.

> *"… we have recently installed solar panels on the roof to power the stairway passage, lifts and water pumps in one of the buildings…the bill of the building maintenance has gone down…"* (R3, Mumbai, India)

### 4.3.3 Policy focus points for energy justice

The topic models and the deep-narrative analysis revealed the latent processes that govern energy culture in poverty. It also revealed a critical nexus between the built environment and energy culture that influences specific human-scale energy services and the poverty penalty associated with it. We argue that these nexuses represent the vertices of distributive energy justice that can aid in designing the required 'good' energy policy. Thus, answering the primary research question of this study. In doing so, we recognise the following policy focus points based on the case studies presented in this study (see Table 4). Each topic represents one policy focus vertices such that a policymaker can include multidimensionality in energy and poverty alleviation policies. For example, topic 1 indicate the quality of the built environment as a crucial distributive justice factor along with affordability and accessibility of appliance ownership (see topic 1 and topic 2, table 4). Similarly, the role of the community-driven initiative is also very high in just policymaking, especially with renewable-based solutions to reduce grid dependencies.

A very crucial just policy variable is energy governance of utilities in poor communities such that they do not push occupants into poverty traps. It is also important to pin-point the causes behind electricity theft and non-uniform metering of houses as a critical measure of distributive justice from the utility side. Such that the utilities can cover their costs somehow and derive an income for investment) and for customers (who are, often understandably, suspicious of the utilities but presumably have a shared interest in reliable services that they can access for a fair price. Besides, provisioning of verified and efficient pre-owned appliances can also contribute to distributive justice through improved well-being (see Table 4). To achieve distributive energy justice and poverty alleviation goals, it is vital to visualise energy use from a multi-dimensional perspective that links sustainability of the built environment, energy systems and human well-being.

Table 4. Derived policy focus points for distributive energy justice and poverty alleviation

| Topics | Policy focus areas for poverty alleviation |
|---|---|
| Topic 1 | Socially inclusive built environment design and planning |
| Topic 2 | Affordable appliance diffusion and social shaping of technology |
| Topic 3 | Community-level renewable energy solutions (rooftop solar PV; biogas; waste-to-energy) |
| Topic 4 | Affordable uptake/easy-loan schemes for welfare appliances, |
| | Trustworthy pre-owned appliance outlets |
| Topic 5 | Socio-economic sensitive energy tariff planning, |
| | Building better governance mechanism of utilities to prevent inconsistent billing, |
| | Transparency in energy billing and metering to customers |
| Topic 6 | Leveraging ICT-based devices and digital platforms for better energy governance, |
| | Consumer complaint infrastructure using ICT-devices, |
| Topic 7 | Adaptive-tariff mechanism based on changing energy culture and appliance uptake |
| | Energy – built environment inclusive banking/financing schemes |

## 5. Concluding remarks

This study forwarded a novel computational social science methodology to perform deep-narrative analysis for energy policy applications. Debnath et al (2020) presented the methodological basis of the proposed nested model, whereas this paper, the part-II illustrates the application of this proposed method. The application was made through a derivation of distributive energy justice pathways in slum rehabilitation built environment of Brazil, India and Nigeria. It contributes to the current literature because such computational policy modelling approach for distributive energy justice was never made before. While the proposed methodology combines artificial intelligence (AI)-driven text mining technique with a qualitative method like grounded theory, the aim was to show the synergistic benefit of using both the method for policy modelling. A growing body of literature is energy research, and social science is garnering narrative and storytelling -based policy modelling methods because of their ability to crystallise arguments. Our methodology further adds to this growing literature by adding statistical robustness to narrative-driven policy modelling methods. In a way, the proposed methodology enables in the identification of 'hotspots' in public deliberations that can have significant impact in evidence-driven policymaking. It contributes to the cutting-edge and multi-disciplinary efforts of using AI-driven tools for public policy applications.

While we automate a preliminary stage of narrative analysis to remove directional biases, we do not intend to replace any of the methods. We merely aim to strengthen the application of computational methods in qualitative analysis to derive generalisable insights, lack of such methods remains a significant literature gap in energy and climate policy research (Ozawa et al., 2019). In doing so, we want to maintain our argument that tools like topic modelling must be used with caution by the analyst to address specific objectives, and not merely as a tool of statistical optimisation. For example, the topics discovered in Fig 9 was reconnected using grounded theory to reveal hotspots of injustices in the energy-culture narratives to derive multidimensional policy focus points in the slum rehabilitation housing.

Using computational tools like topic modelling in a complementary manner with robust qualitative analysis tools will aid in the small-data analysis as we presented here. Otherwise, computational methods only retain their effectiveness with big-data analysis, making the entire process of deep-narrative analysis futile and irrelevant to an experienced qualitative researcher. We also argue that removal of directional bias in the narrative text-data object further enhances the applicability of our method in qualitative policy research and is worth the extra effort of running and validating topic models (like LDA). The capability of the proposed nested-method to retain similarities across heterogenous narratives cases of energy-culture in slum rehabilitation housing of India, Brazil and Nigeria show that while the context may be different, but the problems remain the same. It is therefore crucial to strive for generalisable evidence-based policymaking from micro-data to enable distributive justice.

We would also like to the point that this study is in a preliminary stage, and its effectiveness can only be guaranteed unless an alternative method of distributive justice policy modelling is developed. Removing directionality bias in the primary data binds epistemologically with the philosophy of 'equity and equality' in energy justice concepts, that gives our Bayesian-based method more significant application support. However, the inherent challenges with multi-lingual data collection and transcription remains a constraint of such text-based narrative analysis. And no amount of AI-based data cleaning and deconstruction can address the semantic-biases that may arise during transcription. Our intention was to forward a novel model that can improve the effectiveness of narrative-based policy design methods, such that more effective and appropriate energy and climate policies can be made through public deliberations. Such multi-disciplinary approaches are even more critical today as the traditional technocratic policy modelling realm is reaching its ceiling.

# Acknowledgement

This study is in part supported by the Bill and Melinda Gates Foundation through the Gates Cambridge Scholarship [OPP1144]. Part of this study is also supported by the Santander Mobility Grant for 2018-19 at the University of Cambridge. All opinion, findings and/or conclusion are that of the authors and do not necessarily reflect the view of the funding organisations.

RD would also like to thank the Centre for Urban Science and Engineering at Indian Institute of Technology Bombay for their support. We thank the participants for their time and cooperation.

# References

Aklin, M., Bayer, P., Harish, S. P., & Urpelainen, J. (2018). *Escaping the Energy Poverty Trap: When and How Governments Power the Lives of the Poor*. Boston, USA: The MIT Press.

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Aria, M., & Cuccurullo, C. (2019). biblioshiny-bibliometrix for no coders. Retrieved January 6, 2020, from https://www.bibliometrix.org/biblioshiny/

Arun, R., Suresh, V., Veni Madhavan, C. E., & Murthy Narasimha, M. N. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In M. J. Zaki, J. Xu Yu, B. Ravindran, & V. Pudi (Eds.), *Advances in Knowledge Discovery and Data Mining. PAKDD 2010. Lecture Notes in Computer Science* (pp. 391–402). Hyderabad: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13657-3_43

Bardhan, R., Sarkar, S., Jana, A., & Velaga, N. R. (2015). Mumbai slums since independence: Evaluating the policy outcomes. *Habitat International*, *50*, 1–11. https://doi.org/10.1016/j.habitatint.2015.07.009

Bartiaux, F., Vandeschrick, C., Moezzi, M., & Frogneux, N. (2018). Energy justice, unequal access to affordable warmth, and capability deprivation: A quantitative analysis for Belgium. *Applied Energy*, *225*(March), 1219–1233. https://doi.org/10.1016/j.apenergy.2018.04.113

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(4–5), 993–1022. https://doi.org/10.1016/b978-0-12-411519-4.00006-9

Boarnet, M. G., Giuliano, G., Hou, Y., & Shin, E. J. (2017). First/last mile transit access as an equity planning issue. *Transportation Research Part A: Policy and Practice*, *103*, 296–310. https://doi.org/10.1016/j.tra.2017.06.011

Bombaerts, G., Jenkins, K., Sanusi, Y. A., & Guoyu, W. (Eds.). (2020). *Energy Justice Across Borders* (1st ed.). Springer Open. https://doi.org/10.1007/978-3-030-24021-9

Bouzarovski, S., & Simcock, N. (2017). Spatializing energy justice. *Energy Policy*, *107*(October 2016), 640–648. https://doi.org/10.1016/j.enpol.2017.03.064

Brand-Correa, L. I., Martin-Ortega, J., & Steinberger, J. K. (2018). Human Scale Energy Services: Untangling a 'golden thread.' *Energy Research and Social Science*, *38*(August 2017), 178–187. https://doi.org/10.1016/j.erss.2018.01.008

Brand-Correa, L. I., & Steinberger, J. K. (2017). A Framework for Decoupling Human Need Satisfaction From Energy Use. *Ecological Economics*, *141*, 43–52. https://doi.org/10.1016/j.ecolecon.2017.05.019

Bryant, A., & Charmaz, K. (2019). *The SAGE Handbook of Current Developments in Grounded Theory*. London, UK: SAGE Publications Ltd.

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, *72*(7–9), 1775–1781. https://doi.org/10.1016/j.neucom.2008.06.011

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, *5*(1), 146–166. https://doi.org/10.1016/j.joi.2010.10.002

Day, R., Walker, G., & Simcock, N. (2016). Conceptualising energy use and energy poverty using a capabilities framework. *Energy Policy*, *93*, 255–264. https://doi.org/10.1016/j.enpol.2016.03.019

Debnath, R., Bardhan, R., & Sunikka-Blank, M. (2019a). Discomfort and distress in slum rehabilitation: Investigating a rebound phenomenon using a backcasting approach. *Habitat International*, *87*, 75–90. https://doi.org/10.1016/j.habitatint.2019.03.010

Debnath, R., Bardhan, R., & Sunikka-Blank, M. (2019b). How does slum rehabilitation influence appliance ownership? A structural model of non-income drivers. *Energy Policy*, *132*(December 2018), 418–428. https://doi.org/10.1016/j.enpol.2019.06.005

Fukuyama, H., & Weber, W. L. (2009). A directional slacks-based measure of technical inefficiency. *Socio-Economic Planning Sciences*, *43*(4), 274–287. https://doi.org/10.1016/j.seps.2008.12.001

Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: Interviews and focus groups. *British Dental Journal*, *204*(6), 291–295. https://doi.org/10.1038/bdj.2008.192

Govindan, M., Palit, D., Murali, R., & Sankar, D. (2020a). Gender in Electricity Policymaking in India, Nepal and Kenya. In Gunter Bombaerts, K. Jenkins, Y. A. Sanusi, & W. Guoyu (Eds.), *Energy Justice Across Borders* (pp. 111–135). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-24021-9_6

Govindan, M., Palit, D., Murali, R., & Sankar, D. (2020b). Gender in Electricity Policymaking in India, Nepal and Kenya. In Gunter Bombaerts, K. Jenkins, Y. A. Sanusi, & W. Guoyu (Eds.), *Energy Justice Across Borders* (pp. 111–135). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-24021-9_6

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(SUPPL. 1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Grover, D., & Daniels, B. (2017a). Social equity issues in the distribution of feed-in tariff policy benefits: A cross sectional analysis from England and Wales using spatial census and policy data. *Energy Policy*, *106*(January), 255–265. https://doi.org/10.1016/j.enpol.2017.03.043

Grover, D., & Daniels, B. (2017b). Social equity issues in the distribution of feed-in tariff policy benefits: A cross sectional analysis from England and Wales using spatial census and policy data. *Energy Policy*, *106*(January), 255–265. https://doi.org/10.1016/j.enpol.2017.03.043

Hankerson, D., Harris, G. A., & Peter D Johnson, J. (Eds.). (2003). *Introduction to Information Theory and DataCcompression* (2nd ed.). Washington DC: CRC Press.

Hutt, R. (2016). These are the world's five biggest slums. Retrieved January 9, 2020, from https://www.weforum.org/agenda/2016/10/these-are-the-worlds-five-biggest-slums/

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, *78*(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Jenkins, K., Mccauley, D., Heffron, R., Stephan, H., & Rehner, R. (2016). Energy justice : A conceptual review. *Energy Research and Social Science*, *11*, 174–182. https://doi.org/dx.doi.org/10.1016/j.erss.2015.10.004

Kruger, R., & McCauley, D. (2020). Energy Justice, Hydropower and Grid Systems in the Global South. In G. Bombaerts, K. Jenkins, Y. Sanusi, & W. Guoyu (Eds.), *Energy Justice Across Borders* (pp. 91–109). Cham: Springer Open. https://doi.org/10.1007/978-3-030-24021-9_5

Lee, J., & Byrne, J. (2019). Expanding the conceptual and analytical basis of energy justice: beyond the three-tenet framework. *Frontiers in Energy Research*, *7*(SEP), 1–10. https://doi.org/10.3389/fenrg.2019.00099

Li, Y., Rapkin, B., Atkinson, T. M., Schofield, E., & Bochner, B. H. (2019). Leveraging Latent Dirichlet Allocation in processing free-text personal goals among patients undergoing bladder cancer surgery. *Quality of Life Research*, *28*(6), 1441–1455. https://doi.org/10.1007/s11136-019-02132-w

MacKay, D. J. C. (2008). *Sustainable Energy - Without the Hot Air* (1st ed.). Cambridge, England: UIT Cambridge Ltd.

Manning, C. D., Raghavan, P., & Schutze, H. (2009). *Introduction to Information Retrieval* (1st ed.). Cambridge, England: Cambridge University Press. Retrieved from https://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf

McCauley, D, Heffron, R., Steohan, H., & Jenkins, K. (2013). Advancing energy justice : the triumvirate of tenets and systems thinking. *International Energy Law Review*, *32*(3), 107–116. Retrieved from http://hdl.handle.net/10023/6078

McCauley, Darren, Ramasar, V., Heffron, R. J., Sovacool, B. K., Mebratu, D., & Mundaca, L. (2019). Energy justice in the transition to low carbon energy systems: Exploring key themes in interdisciplinary research. *Applied Energy*, *233–234*(November 2018), 916–921. https://doi.org/10.1016/j.apenergy.2018.10.005

Mitlin, D., & Sattherthwaite, D. (2012). *Urban poverty in the global south: scale and nature* (1st ed.). Oxon: Routledge. Retrieved from https://content.taylorfrancis.com/books/download?dac=C2012-0-00368-4&isbn=9781136249167&format=googlePreviewPdf

Moor, N. (2019). ldatuning: Tuning of the Latent Dirichlet Allocation Model Parameters: R package version 0.2.0. Retrieved November 27, 2019, from https://cran.r-project.org/web/packages/ldatuning/ldatuning.pdf

Munro, P. G., & Schiffer, A. (2019). Ethnographies of electricity scarcity: Mobile phone charging spaces and the recrafting of energy poverty in Africa. *Energy and Buildings*, *188–189*, 175–183. https://doi.org/10.1016/j.enbuild.2019.01.038

Nijman, J. (2008). Against the odds: Slum rehabilitation in neoliberal Mumbai. *Cities*, *25*(2), 73–85. https://doi.org/10.1016/j.cities.2008.01.003

Ozawa, M., Chaplin, J., Pollitt, M., Reiner, D., & Warde, P. (Eds.). (2019). *In Search of Good Energy Policy*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/9781108639439

Robinson, D., & Silge, J. (2019). Package ‘ tidytext .’ Boston, USA: https://cran.r-project.org/. Retrieved from https://cran.r-project.org/web/packages/tidytext/tidytext.pdf

Rondinelli, D. A. (1990). Housing the Urban Poor in Developing Countries:The Magnitude of Housing Deficiencies and the Failure of Conventional Strategies Are World-Wide Problems. *The American Journal of Economics and Sociology*, *49*(2), 153–166. https://doi.org/10.1111/j.1536-7150.1990.tb02269.x

Samarakoon, S. (2019). A justice and wellbeing centered framework for analysing energy poverty in the Global South. *Ecological Economics*, *165*(July), 106385. https://doi.org/10.1016/j.ecolecon.2019.106385

Scartezini, V. (2013). Redeeming e-waste in Brazil. Retrieved January 9, 2020, from https://waste-management-world.com/a/redeeming-e-waste-in-brazil

Shove, E. (2003). *Comfort, cleanliness and convenience : the social organization of normality*. Oxford, UK: BERG.

Shyu, C. W. (2014). Ensuring access to electricity and minimum basic electricity needs as a goal for the post-MDG development agenda after 2015. *Energy for Sustainable Development*, *19*(1), 29–38. https://doi.org/10.1016/j.esd.2013.11.005

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach*. (N. Tache, Ed.) (1st ed.). Sebastopol, CA: O'Reilly Media. Retrieved from https://www.tidytextmining.com/

Silge, J., & Robinson, D. (2019). Introduction to tidytext. Retrieved November 26, 2019, from https://cran.r-project.org/web/packages/tidytext/vignettes/tidytext.html

Sovacool, B. K. (2011). Conceptualizing urban household energy use: Climbing the "Energy Services Ladder." *Energy Policy*, *39*(3), 1659–1668. https://doi.org/10.1016/j.enpol.2010.12.041

Sovacool, B. K., & Dworkin, M. H. (2014). *Global Energy Justice: Principles, Problem and Practices*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CB09781107323605

Sovacool, B. K., & Dworkin, M. H. (2015). Energy justice: Conceptual insights and practical applications. *Applied Energy*, *142*, 435–444. https://doi.org/10.1016/j.apenergy.2015.01.002

Stephenson, J. (2018). Sustainability cultures and energy research: An actor-centred interpretation of cultural theory. *Energy Research and Social Science*, *44*(September 2017), 242–249. https://doi.org/10.1016/j.erss.2018.05.034

Stephenson, J., Barton, B., Carrington, G., Gnoth, D., Lawson, R., & Thorsnes, P. (2010). Energy cultures: A framework for understanding energy behaviours. *Energy Policy*, *38*(10), 6120–6129. https://doi.org/10.1016/j.enpol.2010.05.069

Sunikka-Blank, M., Bardhan, R., & Haque, A. (2019). Gender, domestic energy and design of inclusive low-income habitats: A case of slum rehabilitation housing in Mumbai. *Energy Research and Social Science*, *49*(March), 53–67. https://doi.org/10.1016/j.erss.2018.10.020

Tesfamichael, M., Bastille, C., & Leach, M. (2020). Eager to connect, cautious to consume: An integrated view of the drivers and motivations for electricity consumption among rural households in Kenya. *Energy Research and Social Science*, *63*(November 2019). https://doi.org/10.1016/j.erss.2019.101394

Trotter, R. T. (2012). Qualitative research sample design and sample size: Resolving and unresolved issues and inferential imperatives. *Preventive Medicine*, *55*(5), 398–400. https://doi.org/10.1016/j.ypmed.2012.07.003

Tucho, G. T. (2020a). The Impacts of Policy on Energy Justice in Developing Countries. In Gunter Bombaerts, K. Jenkins, Y. A. Sanusi, & W. Guoyu (Eds.), *Energy Justice Across Borders* (pp. 137–154). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-24021-9_7

Tucho, G. T. (2020b). The Impacts of Policy on Energy Justice in Developing Countries. In Gunter Bombaerts, K. Jenkins, Y. A. Sanusi, & W. Guoyu (Eds.), *Energy Justice Across Borders* (pp. 137–154). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-24021-9_7

UN. (2012). Secretary-general to Global Development Center: 'Energy Is the Golden Thread' Connecting Economic Growth, Social Equity, Environmental Sustainability. Retrieved November 20, 2019, from https://www.un.org/press/en/2012/sgsm14242.doc.htm

UN. (2018). *Advancing SDG 7 implementation in support of the 2030 agenda*. Retrieved from https://sustainabledevelopment.un.org/content/documents/17462PB1.pdf

Walker, G., Simcock, N., & Day, R. (2016). Necessary energy uses and a minimum standard of living in the United Kingdom: Energy justice or escalating expectations? *Energy Research and Social Science*, *18*, 129–138. https://doi.org/10.1016/j.erss.2016.02.007

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag. Retrieved from https://ggplot2.tidyverse.org/

Wickham, H., François, R., Henry, L., & Müller, K. (2019). dplyr: A Grammar of Data Manipulation. Retrieved November 28, 2019, from https://dplyr.tidyverse.org/

Zhou, S., & Noonan, D. S. (2019). Justice implications of clean energy policies and programs in the United States: A theoretical and empirical exploration. *Sustainability (Switzerland)*, *11*(3). https://doi.org/10.3390/su11030807

# Appendix

## A1. Bibliometric topic analysis

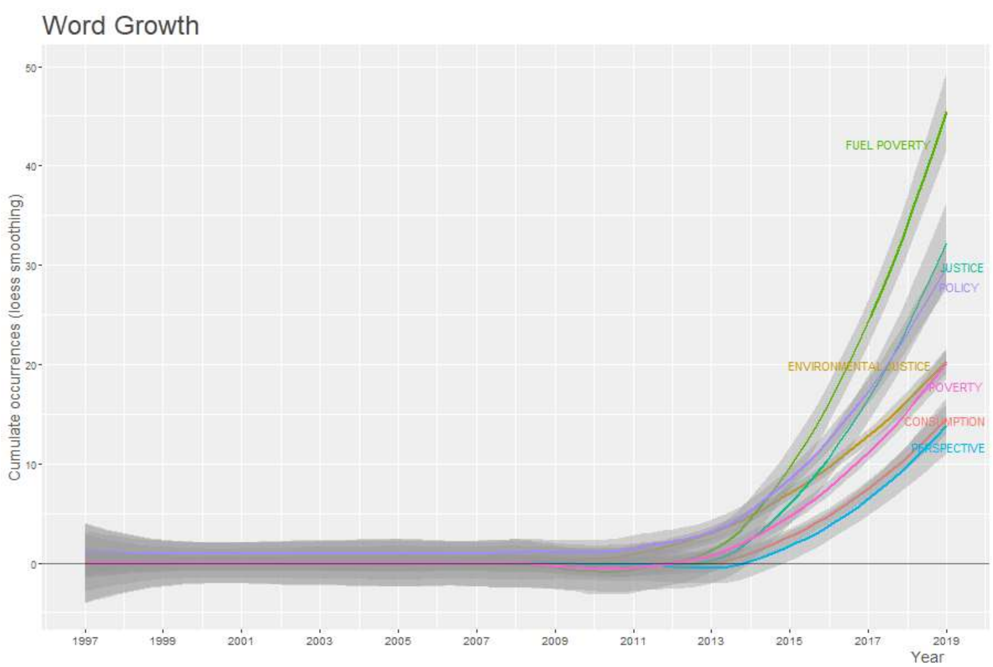



Fig A 1. Topic growth in the published documents (n = 144) on energy justice and poverty between 1997 – 2019.

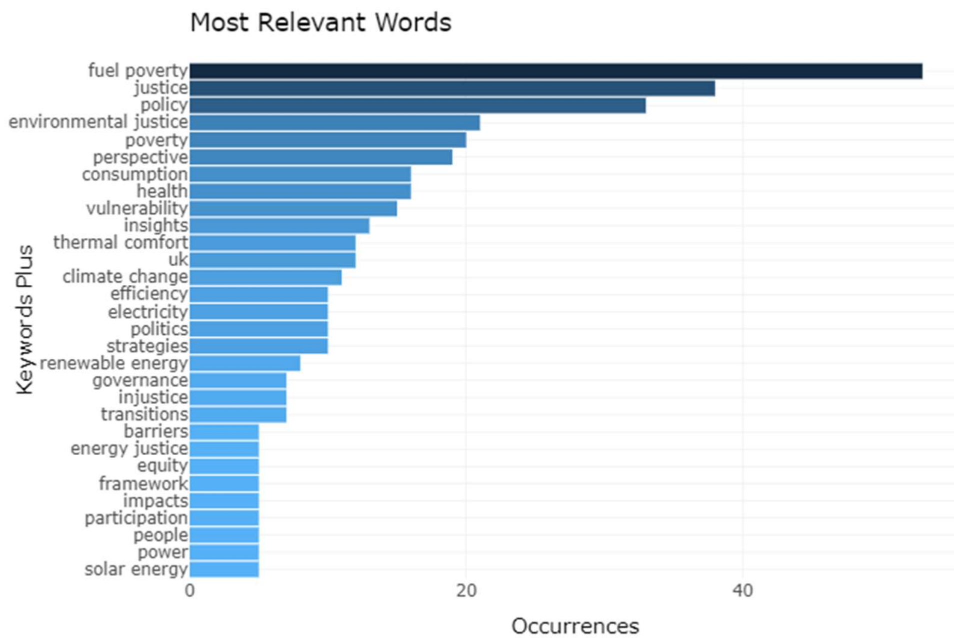



Fig A 2. Occurrence of keywords in the Web of Science database on the topic 'energy justice' and 'poverty'.

## A2 Mathematical background: LDA

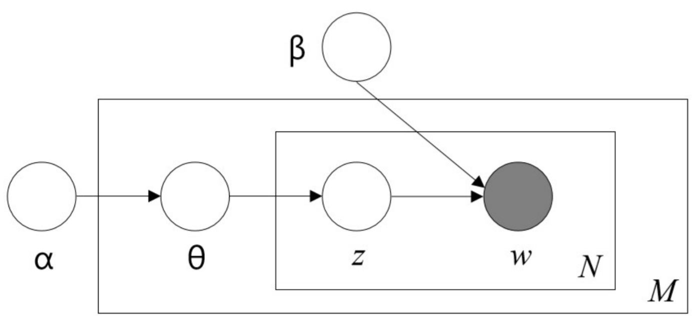



Fig A3. Graphical model representation of LDA. The boxes are "plates" representing replicates. The outer plate represents documents (*M*), while the inner plate represents the repeated choice of topics (*z*) and words (*w*) within a document (*N*). 'Θ' is the topic distribution for document *i*. 'α', 'β' are the hyperparameters of Dirichlet distribution (Source: (Blei, Ng, & Jordan, 2003)).

Topic modelling refers to the task of identifying topics that best describes a set of documents. These topics remain latent and will only emerge during the modelling process. It is based on the basic idea that each document can be expressed as a distribution of topics and each topic can be described by a distribution of words (Blei et al., 2003). The basic terminology used in LDA is based on the language of 'text collection', referring to entities such as "words", "documents" and "corpora". These terms are defined as (Blei et al., 2003):

- A *word* is the basic unit of discrete data, defined to be an item from a vocabulary indexed by $\{1, \ldots, V\}$. We represent words using unit-basis vectors that have a single component equal to one and all other components equal to zero. Thus, using superscripts to denote components, the *v*th word in the vocabulary is represented by a *V*-vector *w* such that $w^v = 1$ and $w^u = 0$ for $u \neq v$.
- A *document* is a sequence of *N* words denoted by $\mathbf{w} = (w_1, w_2, \ldots, w_N)$, where $w_N$ is the *n*th word in the sequence.

- A *corpus* is a collection of M documents denoted by $D = \{\mathbf{w_1}, \mathbf{w_2}, \dots, \mathbf{w_N}\}$.

Latent Dirichlet allocation (LDA) is a generative probabilistic model of a corpus. It is a mixed-membership model which considers that each datapoint may belong to more than a single cluster. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. For the list of assumptions adopted in the basic LDA model, please refer to the Section 3 of Blei et al., (2003). The probability distribution function of a *k*-dimensional Dirichlet random variable θ is illustrated in eq (1), where θ can take values in the (*k*-1)-simplex (a *k*-vector θ lies in the (*k*-1)-simplex if $\theta_i \geq 0, \sum_{i=1}^{k} \theta_i = 1$),

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \dots \theta_k^{\alpha_k - 1}, \tag{1}$$

where the parameter α is a *k*-vector with components $\alpha_i > 0$, and where Γ(x) is the Gamma function. The Dirichlet distribution is a conjugate to the multinomial distribution, and the joint distribution of a topic mixture θ, a set of *N* topics **z**, and a set of *N* words **w**, along with the hyperparameters α and β is given as (see eq (2)),

$$p(\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{w}|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta)\ p(w_n|z_n, \beta), \tag{2}$$

where $p(z_n|\theta)$ is simply $\theta_i$ for the unique *i* such that $z_n^i = 1$. Integrating over θ and summing over *z*, the marginal distribution of the document is obtained (see eq (3)),

$$p(\boldsymbol{w}|\alpha, \beta) = \int p(\theta|\alpha) \left( \prod_{n=1}^{N} \sum_{z_n} p(z_n|\theta)\ p(w_n|z_n, \beta) \right) d\theta \tag{3}$$

The probability of the corpus is obtained by taking the product of marginal probabilities of single documents, and is represented as (see eq (4)),

$$p(D|\alpha, \beta) = \prod_{d=1}^{M} \int p(\theta_d|\alpha) \left( \prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d)\ p(w_{dn}|z_{dn}, \beta) \right) d\theta_d \tag{4}$$

The probabilistic graphical model of LDA is illustrated in Fig 3, which shows three levels to the LDA representation. Hyperparameters α and β are corpus level parameters, assumed to be sampled once in the process of generating a corpus. The variables $\theta_d$ are document-level variables, sampled once per document. Lastly, the variables $z_{dn}$ and $w_{dn}$ are word-level variables and are sampled once for each word in each document (Blei et al., 2003).