

# Using NLP to quantify the environmental cost and diversity benefits of in-person NLP conferences

**Piotr Przybyła**

Institute of Computer Science,  
Polish Academy of Sciences  
Warsaw, Poland

piotr.przybyla@ipipan.waw.pl

**Matthew Shardlow**

Department of Computing and Mathematics,  
Manchester Metropolitan University  
Manchester, UK

m.shardlow@mmu.ac.uk

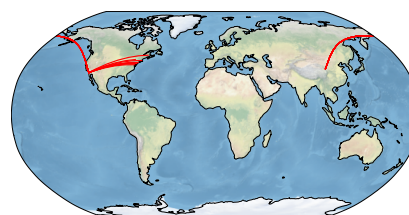
## Abstract

The environmental costs of research are progressively important to the NLP community and their associated challenges are increasingly debated. In this work, we analyse the carbon cost (measured as CO<sub>2</sub>-equivalent) associated with journeys made by researchers attending in-person NLP conferences. We obtain the necessary data by text-mining all publications from the ACL anthology available at the time of the study ( $n=60,572$ ) and extracting information about an author's affiliation, including their address. This allows us to estimate the corresponding carbon cost and compare it to previously known values for training large models. Further, we look at the benefits of in-person conferences by demonstrating that they can increase participation diversity by encouraging attendance from the region surrounding the host country. We show how the trade-off between carbon cost and diversity of an event depends on its location and type. Our aim is to foster further discussion on the best way to address the joint issue of emissions and diversity in the future.

## 1 Introduction

Figure 1 shows the increase in travel to the ACL annual meeting over the past 40 years. Whereas conferences used to be the privilege of a few academics, they are now attended by participants from companies, research institutes and universities across the world. This comes with an increase in the total volume of work published, and with it an increase in the carbon emissions attributed to travelling to in-person events.

In this study we seek to quantify the impact of conferences that are increasingly diverse in terms of participation and location (undoubtedly beneficial) on the increased carbon emissions (undoubtedly detrimental). We base our analysis on publications spanning 55 years (1965–2020), taken from



(a) ACL 1979: La Jolla, California, USA



(b) ACL 1999: College Park, Maryland, USA



(c) ACL 2019: Florence, Italy

Figure 1: Visualisation of estimated journeys to the ACL annual meetings over 40 years. Maps for all major NLP conferences are included in the supplementary material.

the ACL Anthology<sup>1</sup>. We use NLP tools to parse each document and identify the locations of the conference venues and lead researcher's institution. We answer the following questions:

1. Where is NLP research performed and presented?
2. What are the environmental costs?
3. Do conferences increase local participation?
4. Which events attract a diverse audience and how do they compare to non-physical venues?

To the best of our knowledge, our work is the first to quantitatively explore the relationship between the location of conferences in a research field

<sup>1</sup><https://www.aclweb.org/anthology/>

and diversity of participation. We make our dataset and code available<sup>2</sup> to enable further discussion on the costs and benefits of in-person meetings.

## 2 Related work

**Environmental cost of travel and conferences:** It is a well established fact that conferences come with a climate cost (Ciers et al., 2019), which has recently become greater (Pierce et al., 2020). This has led to calls to reduce or cancel the physical academic conference calendar (Johnson et al., 2020; Reay, 2003; Achakulvisut et al., 2020; Jäckle, 2019; Dwyer, 2013).

The scientific discourse has included measuring and quantifying the emissions costs of conferences and the travel associated with them, from specific events (Astudillo and AzariJafari, 2018), to conference series (Neugebauer et al., 2020), or indeed looking at the total emissions of an entire discipline (Waring et al., 2014; Poom et al., 2017).

Travel is not the only cost associated with academic conferences, or research in general, with one PhD accounting for 21.5 tonnes of CO<sub>2</sub>-equivalent emissions (Achten et al., 2013), of which 35% was attributed to conferences. Recent work shows that in France, a typical research lab might dispense 64% of its carbon outputs on conference travel, with the remaining 36% made up mostly of commuting and energy usage (Mariette et al., 2021)

In response to the pandemic, many conferences have moved temporarily online. A meta-analysis of these online conferences showed that a major result of online delivery was a reduction in the registration fee, promoting access (Mubin et al., 2021). Further, online delivery may allay fears of high travel costs (Raby and Madden, 2021) — as is often the case with top-tier conferences. The main barrier to online participation is a perception of reduced social (rather than academic) opportunities (Raby and Madden, 2021), although this may be overcome through facilitating interpersonal meetings, and social discussion (Achakulvisut et al., 2020). It should be noted that whilst travel is unnecessary in virtual conferences, there is still a quantifiable carbon cost due to the infrastructure required (Ong et al., 2012, 2014; Faber, 2021).

Academic conferences are not without their benefits and a clear advantage of in-person conferences rather than online is the perceived value in social interaction (Raby and Madden, 2021). This argu-

ment is strengthened by the observation that citation rates are higher for work presented across longer distances (Chalvatzis and Ormosi, 2020). An important benefit of conferences is providing an opportunity for researchers to interact with peers from diverse cultural, linguistic, demographic and academic backgrounds. This goal is also recognised within the NLP field.<sup>3</sup>

The high climate cost of academic conferences has led to policy considerations (Bossdorf et al., 2010), including the adoption of carbon offsetting programmes for participants (Holden et al., 2017), wise choices of locations to reduce the average journey distance (Wenner et al., 2019) and mandated reporting of climate costs for conferences (Cugniere et al., 2020). Moving towards the adoption of any of these policies would help to begin the mitigation of the environmental impact of academic travel. Similar discussion has already started in computer science conference communities, e.g. ACM (Pierce et al., 2020).

**Environmental cost of ML and NLP research:** In the field of ML and NLP, there has been an increasing trend towards openness in reporting of the emissions associated with AI research (Schwartz et al., 2020), especially that using deep learning (Henderson et al., 2020). Work has also been undertaken to estimate the overall cost of training machine learning (ML) models — taking into account not only the training time, but also the age of the hardware and server location (García-Martín et al., 2019; Lacoste et al., 2019).

There have been a few efforts within our own field of NLP to better understand the impact that modern techniques are having on the environment and specifically to quantify the emissions costs of training ever larger neural networks (Strubell et al., 2019). Benchmarking of NLP systems in terms of their energy consumption is a viable way to better understand the carbon cost of training such a model (Zhou et al., 2020). Taking into account factors such as resource utilisation can give a more accurate picture of the energy consumption of NLP models (Cao et al., 2020).

A recent trend in NLP is to create low-resource models that provide sufficient performance. For example, light transformer models are quicker to train and consequently have a lower carbon footprint (Sanh et al., 2019; Anderson and Gómez-

<sup>2</sup>[https://github.com/piotrmp/nlp\\_geography](https://github.com/piotrmp/nlp_geography)

<sup>3</sup><https://acl2020.org/blog/diversity-and-inclusion/>

Rodríguez, 2020). Transfer learning presents an opportunity for massive carbon savings. If a model can be trained that requires only minimal retraining for various other subtasks, then this prevents further carbon expenditure down the line. Maximising model reusability is a good strategy for reducing carbon emissions (Kocmi and Bojar, 2020).

### 3 Methods

To be able to answer the questions that motivate this work, we need certain data about the research process, in particular regarding the location of researchers’ affiliations and conference venues. Since no such single source of information existed, we decided to combine publicly available resources to create a new dataset containing the information we required. The process we used to create this resource is detailed below:

**Data structure:** A *publication* is an independent piece of research presented to the community as a journal article or a presentation at a conference. For the purposes of this work, each publication is described by: (1) an *identifier*; (2) the first author’s *affiliation* (identified by the domain name in their e-mail address); (3) the *location* of the first author’s affiliation and (4) an *event*, to which the publication is assigned.

An *event* could be a track at a conference, a co-located meeting (e.g. a workshop) or a volume of a journal. It is described by: (1) an *identifier*; (2) a *name* and (3) a *location* – physical place name in case of in-person events or a special tag (@) in case of journals and virtual conferences.

Note that in this model, we always take into account the first author, while in fact one person may attend a conference to present several publications (resulting in less travels) or more than one author may attend to present a single publication (resulting in more travels). Resolving this issue would require conference registration data, which are not publicly available. Further, the address of the primary affiliation does not necessarily match the researcher’s starting location when travelling to a conference.

**Text mining:** In the process of gathering the data we rely on the XML version of the *ACL Anthology* available on GitHub<sup>4</sup> (we used the version from 17.02.2021). From there we obtain the publications (<paper> tag), associated events

(<volume> tag) with titles and locations.

The crucial information missing from the XML structure is the author’s affiliation and their location. This information is mined from the publication text: we download the publication PDF and use *PyMuPDF*<sup>5</sup> to convert it to plain text. Next, we extract the first e-mail domain occurring in text through regular expressions (allowing for the curly brackets notation for account usernames) and treat it as affiliation identifier. Then, we use *spaCy* (Honnibal et al., 2020) to process the text with the `en_core_web_trf` pipeline, based on *RoBERTa* (Liu et al., 2019). Among the text spans recognised by the named entity recogniser as belonging to the category GPE (geopolitical entity), the one occurring first after the first author’s last name is considered their location. Entities occurring close to each other are grouped, so that multi-part names, such as *Cambridge, Massachusetts (USA)*, are located correctly.

Finally, to interpret the location names for affiliations and events, we use the *Geocoding API* of the *Google Maps API*. This allows us to obtain geographical coordinates (longitude and latitude) and country name for each location. We obtain continent information using the `pycountry-convert` Python package.

**Missing data:** The process described above may leave some of the data fields empty. This may be caused by information being omitted in the XML (year or location for events) or PDF files (affiliation address not provided) or imperfect named entity recognition.

In the case of events, we fill the missing data based on co-located events and manual investigation. We also check which of the conferences in 2020 took place as in-person events in the locations advertised. In the case of affiliations, we look at all other publications with the same affiliation and identify the most common location. We assume this location may also be used for the publication in question. Note that some of the PDF files of the oldest publications are based on scanned typescripts. Extracting information from these would require OCR techniques, but this was not attempted within the described work, resulting in a lower coverage of the earliest publications.

**Diversity computation:** To quantify the participation diversity, we use the Gini coefficient  $G$ . While it was originally proposed for assessing in-

<sup>4</sup><https://github.com/acl-org/acl-anthology/tree/master/data/xml>

<sup>5</sup><https://github.com/pymupdf/PyMuPDF>

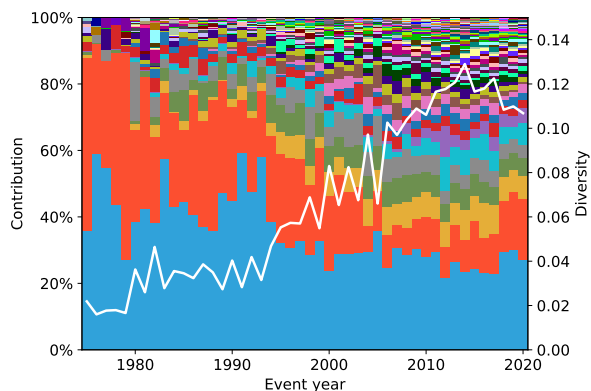


Figure 2: Distribution of NLP publications between affiliation locations (countries) in each year with the diversity index (white line, right axis).

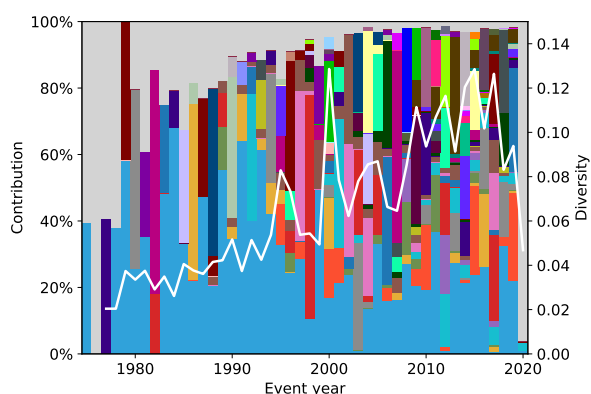


Figure 3: Distribution of NLP publications between event locations (countries, light grey=non-physical venues) with the diversity index (white line, right axis).

come inequality (Gini, 1912), it is widely used as a diversity measure, e.g. of ecosystems (Lexerød and Eid, 2006), research databases (Weidlich and Filippov, 2016) or citation patterns (Leydesdorff et al., 2019). Since  $G$  measures concentration, we define the *diversity* coefficient as  $D = 1.0 - G$ .  $D$  takes values between 0.0 (least uniform distribution, i.e. all conferences happening in the same country) and 1.0 (perfectly uniform distribution, i.e. each country hosting the same number of events).

## 4 Results

The process described above results in a dataset of 60,572 publications associated with 1,991 events. In the following subsections we analyse them to answer some of the important questions about the costs and benefits of the NLP conference system.

**Where is NLP research done?** Regarding affiliations (e-mail domains), we see 5,501 different values in our dataset. Unsurprisingly for literature dating back to 1965, no domain could be found

in a significant portion (22%) of the publications<sup>6</sup>. For the known affiliations, the research output is unequally distributed between them, with the top 207 domains (3.76%) responsible for 50% of the publications. Our diversity index  $D$  takes the value 0.2303.

Regarding addresses, they are associated with 135 countries. Following the refining procedure described in the previous section, only 0.8% unknown values remain. The concentration here is even larger than in the case of affiliations: half of the output is generated by just 3 countries (US, China and Germany) and the  $D$  coefficient equals 0.1087, indicating an even lower diversity amongst international publication in NLP venues.

The contribution varying across years is shown in Figure 2. Coloured bars show the fraction of publications from a given year associated with each country, sorted by their global contribution (US=blue, China=orange, Germany=gold, UK=green, Japan=grey, France=light blue). Additionally, we show the diversity coefficient for the years (white line, right axis). We can see the diversity was rising through most of the considered period, but since 2013, the trend is reversed.

**Where is NLP research presented?** In total, the 1,991 events were held in 48 different countries. The distribution of publications presented at each country is more uniform than previously covered, with diversity index of 0.3838.

Figure 3 shows how this distribution changed across the years. The bars correspond to the number of papers presented in each country in a given year, with the same colour coding as in Figure 2. We can see that the distribution changes drastically every year due to major conferences moving around the world. As previously, we see the increasing diversity through the increasing  $D$  coefficient. Moreover, while the number of articles presented in the most common country (US) was consistently high throughout the studied period, its relative contribution to the overall publication volume was falling for many years. Similarly to the previous plot, a new trend of falling diversity is visible from 2015. Finally, we can observe the changing role of non-physical venues (light grey bars): the share generated by online journals falling over the years and the sudden change in 2020, when 96% of work was presented online.

<sup>6</sup>The oldest available email addresses are located in the .arpa domains.

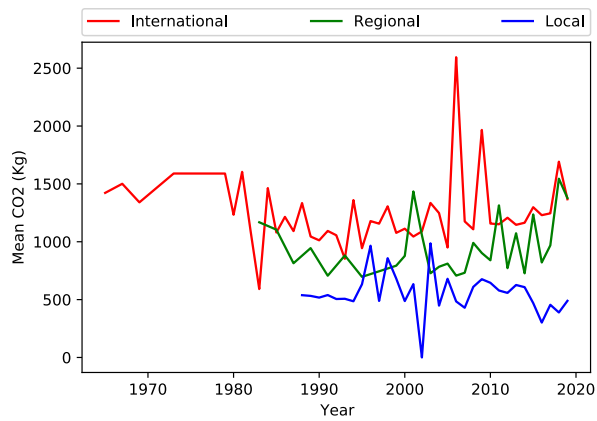


Figure 4: Average emissions per publication at local, regional and international conferences between 1965 and 2019

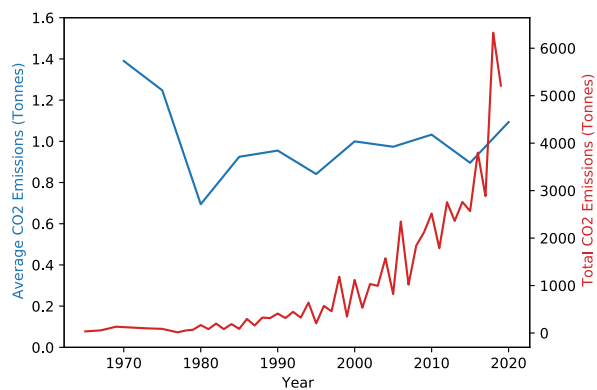


Figure 5: The average emission per publication (over 5-year periods) and total emission (yearly) between 1970 and 2019.

**What are the environmental costs?** Our dataset includes 51,116 publications, for which both the location of research centre and conference venue are known. The average journey distance was 4,988 km and the longest distance travelled was 19,888 km from New Zealand to Spain.<sup>7</sup>

To convert from the number of kilometres travelled (to the conference and back) to the carbon emissions costs, we turned to data from the UK Government for enabling companies to report their emissions<sup>8</sup>. This resource provided us with 5 years of historic emissions data (2016-2020) for short-haul and long-haul flights giving the CO<sub>2</sub> per passenger per kilometer for each given year. We trained a linear regression model to estimate the carbon cost of air travel beyond this time span. Gains in flight efficiency have led to the reduction of carbon emission, resulting in higher costs for his-

<sup>7</sup>Note that this (and other journeys) may have involved a connecting flight, increasing its length.

<sup>8</sup><https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2020>

toric journeys. We used values for CO<sub>2</sub>-Equivalent with Radial Forcing, which give an estimate of the overall climate change impact of travel. We considered international flights as those longer than 3700 km in accordance with the guidelines associated with the data source. Journeys under this were considered short haul, except for those less than 500 km, where we assumed that another lower carbon means of travel would be more likely (in our case we used figures from the same data for train journeys). The data used to create the univariate linear regressions for predicting historic emissions are included in Appendix A.

Each event could be simply represented through its total emissions, but there are several issues with this approach. Firstly, the size of a conference (number of attendees) dictates its overall emissions cost. Therefore, we use the mean carbon cost of a publication at each event instead. Secondly, we compared events according to their geographic reach. International conferences are those that can be hosted anywhere in the world. Regional conferences are those that are restricted to a specific region (we included LREC, which typically happens around the Mediterranean) and local conferences are those that happen in a single country (or a very narrow geographical region). The conferences included in each band are shown in Appendix B.

Figure 4 shows that international and regional conferences are the main emitters of greenhouse gasses in the NLP field. Local conferences emit around a quarter of the CO<sub>2</sub>-Equivalent (per publication) compared to international or regional conferences. Whilst regional conferences have traditionally tracked below the average emissions of international conferences, the gap between them is narrowing, as these conferences are increasingly treated as international events.

Figure 5 shows the discrepancy between the total CO<sub>2</sub> emissions (in red, right axis) and the average CO<sub>2</sub> emissions (in blue, left axis) over the same period across our entire dataset. We can see that whilst the average emissions fluctuate, they are generally stable around 0.8-1.2 tonnes of CO<sub>2</sub> emitted per publication. This stability is possibly due to the fact that the increasing distances travelled are offset by increasing flight efficiency. In contrast, the total amount of CO<sub>2</sub>-equivalent emitted by conferences has risen exponentially hitting 1 million kg in 1998, 2 million kg in 2006, 3 million kg in 2016 and then jumping to over 6 million kg in 2018.

| 6,000 Tonnes of CO <sub>2</sub> -Equivalent equates to... |   |
|---|---|
| 1,304   | cars driven for a year                      |
| 722   | homes powered for a year                    |
| 13,892  | barrels of oil (energy production)          |
| 99,212  | new trees planted (CO <sub>2</sub> capture) |
| 339,172   | NLP pipelines trained                       |
| 168   | NLP pipelines optimised                     |
| 68,894  | Generic Transformers trained                |
| 22  | Generic Transformers optimised              |
| 71  | Instances of GPT-3 trained                  |

Table 1: Comparisons of recent annual conference emissions to familiar scenarios both within and outside of NLP.

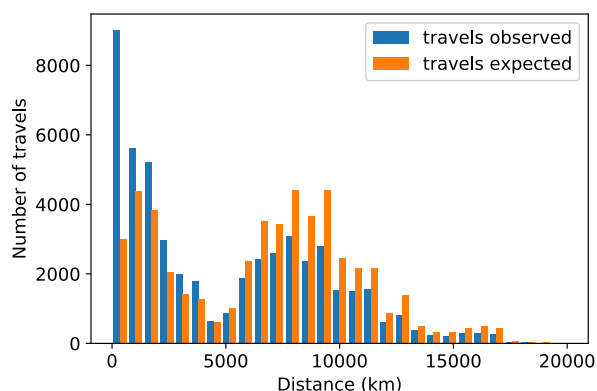


Figure 6: Comparison of the number of travels of certain distance (X axis, in km.) made in two scenarios: observed in the data and expected in case of random choice of events.

To put the value of around 6,000 tonnes of CO<sub>2</sub>-equivalent (total emissions of NLP conferences in 2018) into context, we can compare to emissions for other activities. These are shown in Table 1 and were calculated using data from the website of the US Environmental Protection Agency<sup>9</sup>. Data estimating the amount of emissions used to train NLP models (Strubell et al., 2019; Lasse et al., 2020) are also included.

**What are the diversity benefits?** We hypothesise that series of events occurring in different locations have the benefit of encouraging local researchers to attend, increasing the diversity of participation. In this section we seek to quantify this effect.

Firstly, we verify this hypothesis by comparing the distances researchers travelled for conferences (blue bars) to the distances they would need to travel if they were choosing venues randomly (orange bars) in Figure 6. The results clearly confirm

<sup>9</sup><https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator>

our assumptions: the number of observed short trips, especially a few hundred kilometers, is much higher than expected in a random choice scenario. The number of long trips, especially around 10,000 km, is greatly reduced. Using the data from the previous section, we can also estimate that thanks to these choices, the carbon cost of all travels was 27.21% lower (a total saving of 19,104 tons of CO<sub>2</sub> according to emission rates of 2020).

Next, we can ask whether the priority given to local conferences depends on what country a researcher comes from. To that end, we compute the relative travel length by dividing the observed mean travel distance by the travel distance in a 'random choice' scenario. Figure 7 shows all countries with at least 15 publications according to their relative travel length and GDP per capita in 2018 (Bolt and van Zanden, 2020). We can see that the longest travels are made by countries in the middle-east, most of them considerably wealthy. Most countries that prefer nearby conferences have relatively low income, e.g. Serbia, Philippines or Bulgaria.

Knowing that each event generates diversity by encouraging researchers from the nearby countries to participate, we can now measure how well this effect works for different conferences. It might be expected that achieving high diversity comes at a cost of longer journeys. We verify this by plotting the diversity of in-person events against travel distance (average per publication) in Figure 8. Most events are indeed arranged along an upward direction, but some do not belong to that trend. For example, we can see that EACL conferences deliver more diversity than others for the same travel distance. Some ACL meetings<sup>10</sup>, on the contrary, are associated with very long travel and not so much diversity. LREC events are clear outliers here, since they have by far the highest diversity for low distances. The dashed line corresponding to the diversity index of journals indicates that the diversity observed in many in-person events is much higher. Note that the online conferences are not included in this analysis, since their format was often unclear to authors in the moment of submission.

In Figure 9 we compare the mean participation diversity of events organised in a given continent across the years. Consistently with Figure 2, we see an increasing diversity throughout most of the considered period for most continents. Europe is

<sup>10</sup>COLING/ACL in Sydney (2006), EMNLP/ACL in Singapore (2009) and ACL in Melbourne (2018).

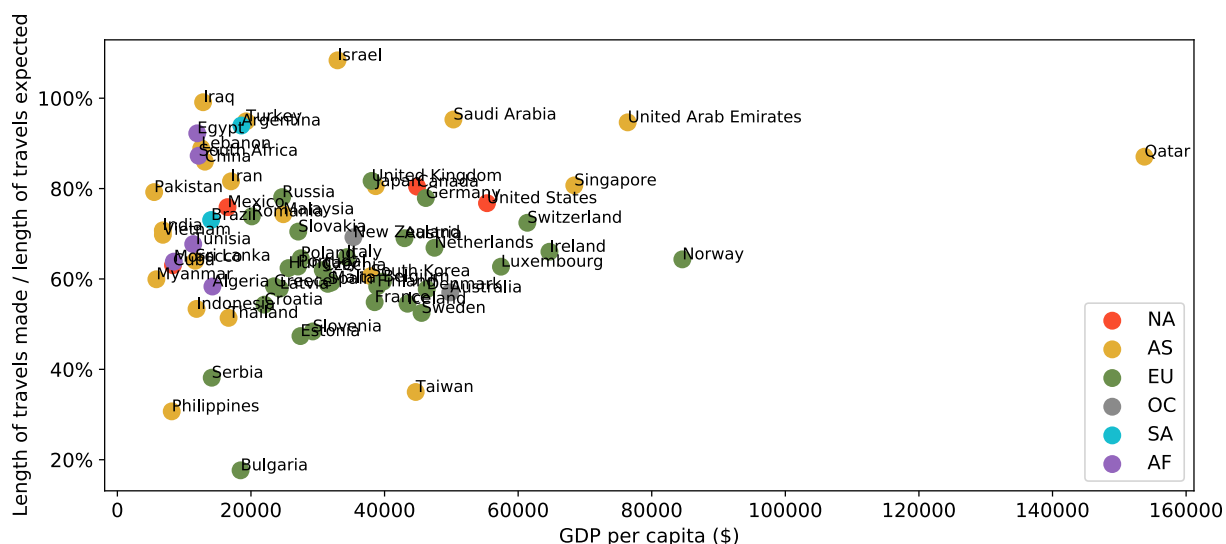


Figure 7: Relative travel length (mean distance of travels made divided by mean distance of travels expected in random venue choice) for countries with at least 15 publications with respect to their continent and GDP per capita.

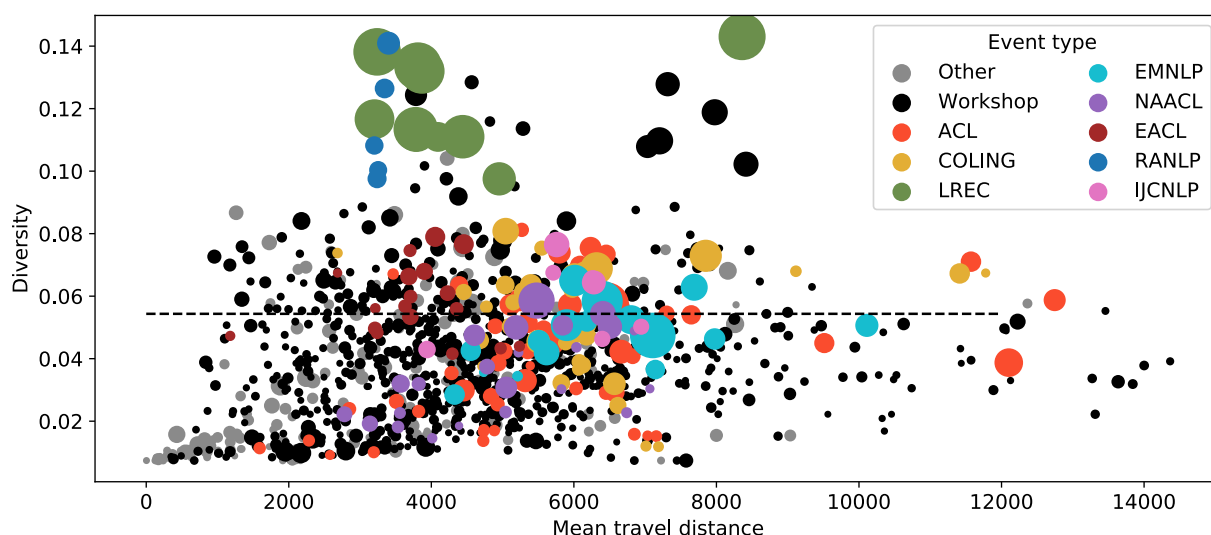


Figure 8: NLP events plotted with respect to the diversity of participation (Y axis), mean travel distance (X axis) and number of publications (disc size).

the location of very diverse events, but the Asian ones appear to be catching up. The journals have seen relatively slow growth and remain much less diverse than in-person events, except for South America or Australia and Oceania, where too few conferences took place for our analysis.

## 5 Discussion

Our work covers the carbon cost and diversity gain associated with conferences in the ACL Anthology. We consider that it is timely to perform this analysis, given the shutdown in physical meetings brought on by the global COVID-19 pandemic and have focussed our analysis on conferences from before the pandemic began.

We have made a number of assumptions in our

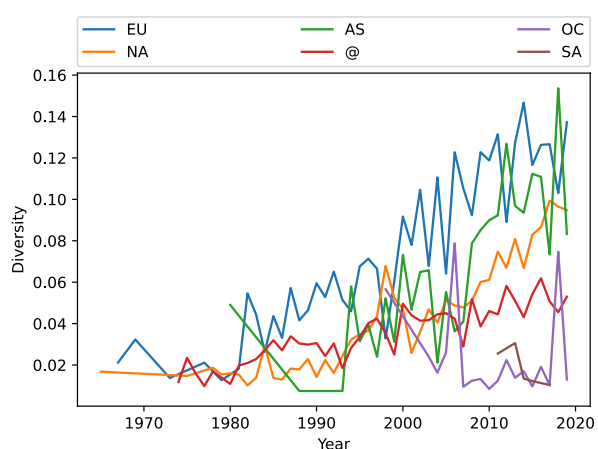


Figure 9: Diversity of events held on each continent between 1965 and 2019. '@' refers to journals. Africa is not represented due to the lack of events there in the ACL Anthology.

work. Most notably, we have assumed that only first authors travel from the location of their institution to the location of the conference (and back) without detour via the easiest means of transport available to them. Our assumptions are consistent between events and as such, our methodology gives a useful tool for comparing potential climate impact in the field of NLP and beyond.

Figure 2 shows that whilst the diversity index has grown consistently from 1970 to 2014, it has dropped since then, with 2020 having the lowest diversity index since 2008. We cannot give an explanation for the drop over this period without speculating, however tracking this index will allow us to measure the change in diversity over the coming years.

Whereas previous work has claimed that non-physical venues promote diversity (Raby and Maden, 2021), our research broadens the picture, with Figure 8 demonstrating that whilst some events are below the mean diversity index of online journals, many are above; in particular LREC and RANLP attract an audience from many countries. We chose not to make a direct comparison between in-person events and the pandemic-era online conferences of 2020 and 2021, since some events of the latter type were (at the point of submission) advertised as physical meetings, while others were in the hybrid format. However, extending our analysis to pure online and hybrid events is a clear direction for future work.

We were also able to quantify the carbon cost of travelling to physical events in terms of CO<sub>2</sub>-equivalent. Whilst this has unsurprisingly grown with the growth of the NLP field, the average carbon cost per paper has remained stable, indicating that gains in efficiency from better modes of transport are offset by an increased travel distance. The total emissions in recent years has been as high as 6,000 tonnes of CO<sub>2</sub>-equivalent. It must also be noted that other activities of NLP research contribute to the total carbon cost generated by the NLP field. For example, the carbon cost of all travel in a single year of NLP research equates to about 22 fully optimised transformer models trained from scratch (see Table 1). We must also address the carbon cost of research, as well as considering the cost of flying to conferences.

Measuring the diversity impact contributed by a conference happening in a certain place is not possible directly, since we cannot know, who would

participate, if the event took place elsewhere. However, our data indicate a preference for local events, which is the highest in low-income countries. Holding conferences across the globe allows researchers from diverse locations to attend an event without flying as far as in a scenario where all conferences were located in one region (as was the case in the early days of the ACL conferences). However, there is a cautionary tale to tell in our data relating to the year 2018. In Figure 5, a large spike on the right hand side corresponds to 2018, when a total of over 6 thousand tonnes of CO<sub>2</sub>-equivalent was attributed to conference travel. In this year the ACL annual meeting was held in Melbourne, Australia and LREC was held in Miyazaki, Japan. The effect of this is clear as researchers from Europe and North America — who usually attend these conferences — needed to travel further, increasing the emissions. Holding conferences in different locations will only lead to increased diversity if these events are advertised to and attended by a majority of people from the region they are held in.

Our definition of diversity index only takes into account the countries from which authors have attended, and does not measure other important factors of diversity (gender, race, economic status, native language, etc.). Whilst some of this information may be discernable from our data, most of it would only be possible to discover by author disclosure, which was not possible in our context. Reporting on the country-based diversity allows us to better understand the diversity of NLP research across the last 50 years.

Our work is designed as a focussed study on the ACL anthology, and a similar analysis of a broader scope (e.g., all computer science, all science publications) would yield results allowing comparisons between disciplines. We were able to perform this analysis due to the provision of the ACL Anthology, which only covers papers in our field. Whilst other resources indexing AI and wider computer science, or even generic scientific literature, do exist (e.g., DBLP, Google Scholar, repositories such as OpenAire, event websites etc.), these each have their own limitations, such as not including PDF links (only DOIs which point to journal websites), lack of a public API or covering only a subset of the literature. Event websites are a fruitful source for data mining, but each event has its own bespoke format and extracting data this way is slow.

We have attempted to give a view of the data that



allows policy makers to make informed decisions on where the next NLP conference should be. We have also made our data available to facilitate future research. Policy makers may wish to consider the high emissions impact of locating a conference in an area far away from the typical attendance base, and also weigh this against the potential diversity gain of locating a conference in a lower-wealth area. We expect that conference organisers will make different decisions based on the relative importance of the above factors to their communities.

## Acknowledgements

This work was supported by the *Polish National Agency for Academic Exchange* through a *Polish Returns* grant number PPN/PPO/2018/1/00006.

## References

- Titipat Achakulvisut, Tulakan Ruangrong, Isil Bilgin, Sofie Van Den Bossche, Brad Wyble, Dan FM Goodman, and Konrad P Kording. 2020. Point of view: Improving on legacy conferences by moving online. *Elife*, 9.
- Wouter MJ Achten, Joana Almeida, and Bart Muys. 2013. Carbon footprint of science: More than flying. *Ecological indicators*, 34:352–355.
- Mark Anderson and Carlos Gómez-Rodríguez. 2020. Distilling neural networks for greener and faster dependency parsing. In *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 2–13.
- Miguel F Astudillo and Hessam AzariJafari. 2018. Estimating the global warming emissions of the Icaxvii conference: connecting flights matter. *The International Journal of Life Cycle Assessment*, 23(7):1512–1516.
- Jutta Bolt and Jan Luiten van Zanden. 2020. [Maddison style estimates of the evolution of the world economy. A new 2020 update](#). Technical report, Maddison Project.
- Oliver Bossdorf, Madalin Parepa, and Markus Fischer. 2010. Climate-neutral ecology conferences: just do it! *Trends in ecology & evolution*, 25(2):61.
- Qingqing Cao, Aruna Balasubramanian, and Niranjan Balasubramanian. 2020. Towards accurate and reliable energy measurement of nlp models. In *Proceedings of SustaiNLP: Workshop on Simple and Efficient Natural Language Processing*, pages 141–148.
- Konstantinos Chalvatzis and Peter L Ormosi. 2020. The carbon impact of flying to economics conferences: is flying more associated with more citations? *Journal of Sustainable Tourism*, 29(1):40–67.
- Joachim Ciers, Aleksandra Mandic, Laszlo Daniel Toth, and Giel Op't Veld. 2019. Carbon footprint of academic air travel: A case study in switzerland. *Sustainability*, 11(1):80.
- Laure Cugniere, Diogo Veríssimo, Angeles Branas, and Guy Bigwood. 2020. [From call to action: a roadmap to sustainable conferences](#). *SocArXiv*.
- James Dwyer. 2013. On flying to ethics conferences: Climate change and moral responsiveness. *IJFAB: International Journal of Feminist Approaches to Bioethics*, 6(1):1–18.
- Grant Faber. 2021. [A framework to estimate emissions from virtual conferences](#). *International Journal of Environmental Studies*, 78(4):608–623.
- Eva García-Martín, Crefeda Faviola Rodrigues, Graham Riley, and Håkan Grahn. 2019. Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing*, 134:75–88.
- Corrado Gini. 1912. *Variabilità e mutabilità*. Rome: Libreria Eredi Virgilio Veschi.
- Peter Henderson, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. 2020. Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248):1–43.
- Matthew H Holden, Nathalie Butt, Alienor Chauvenet, Michaela Plein, Martin Stringer, and Iadine Chadès. 2017. Academic conferences urgently need environmental policies. *Nature ecology & evolution*, 1(9):1211–1212.
- Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. [spaCy: Industrial-strength Natural Language Processing in Python](#).
- Sebastian Jäckle. 2019. We have to change! the carbon footprint of ecpr general conferences and ways to reduce it. *European Political Science*, 18(4):630–650.
- Ruth Johnson, Andrada Fiscutean, and Serghei Mangul. 2020. Refining the conference experience for junior scientists in the wake of climate change. *arXiv preprint arXiv:2002.12268*.
- Tom Kocmi and Ondřej Bojar. 2020. Efficiently reusing old models across languages via transfer learning. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 19–28, Lisboa, Portugal. European Association for Machine Translation.
- Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. 2019. Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*.

- F. Wolff Anthony Lasse, Benjamin Kanding, and Raghavendra Selvan. 2020. Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. In *Proceedings of the ICML Workshop on Challenges in Deploying and monitoring Machine Learning Systems*. ICML.
- Nils L. Lexerød and Tron Eid. 2006. An evaluation of different diameter diversity indices based on criteria related to forest management planning. *Forest Ecology and Management*, 222(1-3):17–28.
- Loet Leydesdorff, Caroline S. Wagner, and Lutz Bornmann. 2019. Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics*, 13(1):255–269.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:11907.11692*.
- Jérôme Mariette, Odile Blanchard, Olivier Berné, and Tamara Ben Ari. 2021. An open-source tool to assess the carbon footprint of research. *arXiv preprint arXiv:2101.10124*.
- Omar Mubin, Fady Alnajjar, Abdullah Shamaail, Suleman Shahid, and Simeon Simoff. 2021. The new norm: Computer science conferences respond to covid-19. *Scientometrics*, 126(2):1813–1827.
- Sabrina Neugebauer, Maren Bolz, Rose Mankaa, and Marzia Traverso. 2020. How sustainable are sustainability conferences?—comprehensive life cycle assessment of an international conference series in europe. *Journal of cleaner production*, 242:118516.
- Dennis Ong, Tim Moors, and Vijay Sivaraman. 2012. Complete life-cycle assessment of the energy/co2 costs of videoconferencing vs face-to-face meetings. In *2012 IEEE Online Conference on Green Communications (GreenCom)*, pages 50–55. IEEE.
- Dennis Ong, Tim Moors, and Vijay Sivaraman. 2014. Comparison of the energy, carbon and time costs of videoconferencing and in-person meetings. *Computer communications*, 50:86–94.
- Benjamin C Pierce, Michael Hicks, Crista Lopes, and Jens Palsberg. 2020. Conferences in an era of expensive carbon. *Communications of the ACM*, 63(3):35–37.
- Age Poom, Kati Orru, and Rein Ahas. 2017. The carbon footprint of business travel in the knowledge-intensive service sector. *Transportation Research Part D: Transport and Environment*, 50:292–304.
- Cassandra L Raby and Joah R Madden. 2021. Moving academic conferences online: Aids and barriers to delegate participation. *Ecology and Evolution*, 11(8):3646–3655.
- David S Reay. 2003. Virtual solution to carbon cost of conferences. *Nature*, 424(6946):251–251.
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.
- Roy Schwartz, Jesse Dodge, Noah A. Smith, and Oren Etzioni. 2020. Green ai. *Communications of the ACM*, 63(12):54–63.
- Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and policy considerations for deep learning in nlp. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650.
- Timothy Waring, Mario Teisl, Eva Manandhar, and Mark Anderson. 2014. On the travel emissions of sustainability science research. *Sustainability*, 6(5):2718–2735.
- Iwona E. Weidlich and Igor V. Filippov. 2016. Using the gini coefficient to measure the chemical diversity of small-molecule libraries. *Journal of Computational Chemistry*, 37(22):2091–2097.
- Fabian Wenner, Freke Caset, and Bart De Wit. 2019. Conference locations and sustainability aspirations: Towards an integrative framework? *disP-The Planning Review*, 55(1):34–51.
- Xiyu Zhou, Zhiyu Chen, Xiaoyong Jin, and William Yang Wang. 2020. Hulk: An energy efficiency benchmark platform for responsible natural language processing. *arXiv preprint arXiv:2002.05829*.

## A Values used in Calculations of Emissions per Passenger

Table 2 shows the Kg of CO<sub>2</sub>-equivalent per passenger used in our calculations to train a univariate linear regression model for historic prediction.

## B Conferences Analysed

To produce Figure 4, we selected specific conferences that we denoted as either local, regional or international. Conferences were selected if they had a specific identifier in the ACL Anthology. The pythonic regular expressions used to match the identifiers and the categorisation of each conference is provided in Table 3. We also used these identifiers to produce the table of travel maps in the supplementary material.

| Mode of Transport | 2020    | 2019    | 2018    | 2017    | 2016    |
|-------------------|---------|---------|---------|---------|---------|
| Long-Haul Flight  | 0.09994 | 0.10244 | 0.11131 | 0.1034  | 0.10035 |
| Short-Haul Flight | 0.08145 | 0.08291 | 0.08503 | 0.08432 | 0.08821 |
| Train Journey     | 0.03659 | 0.04077 | 0.04383 | 0.04636 | —       |

Table 2: Carbon cost (kg of CO<sub>2</sub>-equivalent per passenger) with respect to mode of transport and year.

| Event Name | ACL Anthology Identifiers                                       | Categorisation |
|------------|---|----------------|
| ACL        | r"P\d\d\.\d", r"2020\acl\.main"                                 | International  |
| EMNLP      | r"D\d\d\.[123]", r"2020\emnlp\.main"                            | International  |
| COLING     | r"C\d\d\.\d", r"2020\coling\.main"                              | International  |
| CoNLL      | r"K\d\d\.\d", r"2020\conll\1"                                   | International  |
| NAACL      | r"N\d\d\.\d"  | Regional       |
| LREC       | r"L\d\d\.\d", r"2020\lrec\1"                                    | Regional       |
| EACL       | r"E\d\d\.\d"  | Regional       |
| IJCNLP     | r"I\d\d\.\d", "P15", "D19"                                      | Regional       |
| TALN       | r"F\d\d\.\d", "\d\d\d\d\jeptalnrecital\.*"                      | Local          |
| RANLP      | r"R\d\d\.\d"  | Local          |
| ALTA       | r"U\d\d\.\d"  | Local          |
| PACLIC     | r"Y\d\d\.\d"  | Local          |
| ROCLING    | r"O\d\d\.\d"  | Local          |
| NoDaLiDa   | r"W11\.46", r"W13\.56", r"W15\.18",<br>r"W17\.2\\$", r"W19\.61" | Local          |

Table 3: Regular expressions used to match conferences.