Community Learning: Understanding A Community Through NLP for Positive Impact

Md Towhidul Absar Chowdhury and Naveen Sharma
{mac9908, nxsvse}@rit.edu
Rochester Institute of Technology
Rochester, New York, USA

Abstract
A post-pandemic world resulted in economic upheaval, particularly for the cities’ communities. While significant work in NLP4PI focuses on national and international events, there is a gap in bringing such state-of-the-art methods into the community development field. In order to help with community development, we must learn about the communities we develop. To that end, we propose the task of community learning as a computational task of extracting natural language data about the community, transforming and loading it into a suitable knowledge graph structure for further downstream applications. We study two particular cases of homelessness and education in showing the visualization capabilities of a knowledge graph, and also discuss other usefulness such a model can provide.

1 Introduction
A computational approach to achieving sustainable development goals (SDGs) is of particular interest in the AI for social good domain. There is a significant potential in AI acting as an enabler towards achieving SDGs to benefit society (Vinueza et al., 2020). Out of the 17 SDGs, SDG 11 (Sustainable Cities and Communities) focuses on building communities that are resilient and safe. This includes enabling better infrastructure, building social cohesion, and ensuring a sustainable environment for the growth of a community (Desa and others, 2016). Significant work in NLP for Positive Impact (NLP4PI) has focused on big picture issues including the field of political science (Jin et al., 2021), criminal law (Hong et al., 2021), hope speech (Palakodety et al., 2019), supportive content (Yoo et al., 2021), and national events (Dutta et al., 2022).

A common trend among these research is the focus on national and international events. But there is a gap in research on studying NLP methods and applications to perform data-driven analysis of a city. Cities can be defined as a system of sub-systems - such building blocks are commonly known as communities. Developing such communities, to benefit the city as a whole, have been the focus in social science literature pertaining to community development (Donaldson and Daughtery, 2011; Lalot et al., 2021; Frumkin, 2014; Howard and Johansson, 2014). Community residents are always striving to develop their community, and in this age even at the community level “big data” challenges exist (Chowdhury and Sharma, 2022; Goldsmith and Crawford, 2014).

Before we can focus on an NLP based approach to community development - we need to have a systematic definition of what a community is. A community is modeled as a system of assets that interact with each other to generate capital (value). Community development focuses on the maximization of this capital (Callaghan and Colton, 2008). Each of these assets is a sub-system of its own and can be modeled with its own ontology. Furthermore, these assets interact with each other to provide value to the community i.e. community capital. A taxonomy of the value within the community is already defined in (Callaghan and Colton, 2008) with six major concepts that generate capital - (1) environment (e), (2) human (h), (3) commercial (c), and (4) public (p), (5) cultural, and (6) social.

The first four concepts are easy to define as they have physical assets that can be modeled within the community. For example, trees and parks provide environmental capital. A community’s people, libraries, and local stores represent a human, public and commercial entity respectively. The cultural and social concepts define a set of relationships that have developed around shared values, norms and trust (Callaghan and Colton, 2008). These are the contextual information that is not present in statistical and demographic data because a community
has its people and its unique culture. Any computational model for a community needs to have a structure that captures the context of these people and the culture. Big data approaches that exist in the present literature have not been used to capture this contextual information that exists within the community.

To capture such contextual information, NLP models need to extract this context from raw data. Each time we want to analyze a particular community, we would need to reinvent the wheel - extract the raw data, transform the data into a format for language models to use, and load the data into NLP models.

This is particularly challenging when we attempt to understand a community through a conversational lens. Social media discourse of a community’s challenges and strengths follow jargon that require a computational model to be retrained for each downstream tasks. Establishing a baseline model and a standard pipeline for attempting the analysis of such conversations would go a long way to benefit NLP tasks geared towards community development. Present SOTA methods for establishing standard knowledge models use knowledge graphs for such purposes. Knowledge graphs have been deployed in literature to analyze eco-industrial parks (Lim et al., 2021), healthcare misinformation (Cui et al., 2020), and study social networks (Alonso et al., 2019) with each domain having its own challenges. The benefit of a knowledge graph in all these problems is the ability to not only encapsulate heterogeneous data but also have a schema that is readily available and usable for downstream AI tasks (Ji et al., 2020). Social scientists and community residents can also benefit from the explainable nature of knowledge graphs as they are easy to visualize and human-readable (Van der Maaten and Hinton, 2008; Ji et al., 2020).

In summary, the first step towards utilizing NLP to positively develop a community requires an understanding of the said community. In order to do that, we need to establish a baseline model of storing the knowledge regarding the community. There is a gap in present research in establishing this model for further downstream NLP tasks.

In this paper, we propose the novel task of community learning to map natural language data regarding the community to a knowledge base. We also establish a novel community knowledge graph as the knowledge base for community learning and approach this task as an extract-transform-load (ETL) problem for natural language data.

Our contributions in this paper are the following:

1. Propose the novel task of community learning.
2. Develop a community knowledge graph to encapsulate community data for easier downstream NLP tasks.
3. Introduce our novel ETL pipeline for the automatic construction of a community knowledge graph from raw conversational data.

2 Community Knowledge Graph (CKG)

Let us consider the task of community learning from community data. We define community data as any information (natural language or structured data) that includes information of value for a community. We define a community as a collection of individuals sharing a common characteristic - culture, shared goal, domain interest etc. In our case, with a focus on neighborhoods and cities, we have chosen geographic location as the determinant of a community. For the purpose of this paper, we particularly focus on natural language data and information pertaining to the community capital grounded approach (Callaghan and Colton, 2008) for information of value as stated in Section 1.

We formally define the task of community learning as follows. Given data D about a community, perform a task of community learning CL(D) to return a community knowledge graph CKG. We approach the function CL(D) as a series of extract E, transform T, and load L tasks in mapping community data to a CKG.

For this paper, we limit community data D to only natural language data T to study the richness and importance of contextual conversational data regarding a community. The natural language data can include any textual information from different sources such as forums, wikis, social media posts, and conversations.

2.1 CKG Schema

We define the schema of our community knowledge graph (CKG) as follows:

\[ CKG = \{ \varepsilon, R, F \} \]

where \( \varepsilon \) is the set of entities, \( R \) is the set of relations, and \( F \) is the set of facts in our CKG. The set of entities is defined as:

\[ \varepsilon = Assets \cup OLE \cup Knowledge \]
\[ \text{Assets} = \{e,h,c,p\}, (s,r,o) \in F \]

\[ s, o \in \varepsilon, r \in R \]

where Assets is the set of assets defined in the taxonomy of a community (Callaghan and Colton, 2008) representing environment (e), human (h), commercial (c), and public (p). OLE refers to additional entities we introduced for completeness. The additional concepts in OLE stands for object, event, and location adapted from the POLE data model (Falzone et al., 2021) used for geospatial and temporal analysis. This allows our CKG to store data from multiple modality without needing significant redesign in the future.

The actual entities in the community will be mapped in the set Knowledge. We do not restrict this set to a specific ontology, and entities from this set can be mapped to either Assets or OLE sets with an IS_A relationship. This approach allows us to both capture concepts and implementations of the concepts in our knowledge graph. Furthermore, the social and cultural aspect of our CKG cannot be restricted to a specific ontology in order to ensure it is able to learn from the community using language learning paradigms (Manning et al., 2014; Mitchell et al., 2018) instead of being limited by a pre-determined ontology. It is this social and cultural information that can provide context within the community. These two capitals can be modeled using conversations of community residents (Chowdhury and Sharma, 2021), and mapping social interactions and cohesion (Lalot et al., 2021).

Figure 1 displays the conceptual model of our community knowledge graph, indicating how concepts and actual entities are linked to model a dynamic community.

2.2 Dataset

The data source for our conversational data is the neighborhood discussion social platform Nextdoor (Lambright, 2019). Nextdoor, by default, limits the accessibility of posts from distant neighborhoods and as a result we only parsed conversations in neighborhoods and communities close to our institution (Kurwa, 2019; Masden et al., 2014). Due to anonymity requirements, we are not disclosing the target city and neighborhoods. As per Nextdoor policy we also cannot make our dataset publicly available due to the authentication requirement from our research team. But we will release our code for building and parsing said dataset for use by the research community to build their own conversational dataset.

We built our dataset with search words related to our case studies in Section 4. We focused on specific search words and scraped the posts and conversations that came relevant in our query. Figure 2 shows sample posts for two of the search words we have used. We initially collected over 180,000 posts and replies from the website but finally downsampled to only relevant 28,000 conversations relevant to our two topics of interest.

3 Community Learning

The function \( CL(D) \) is a series of extract \( E \), transform \( T \), and load \( L \) functions on conversational data \( D \) to generate CKG triples \((s,r,o)\).

3.1 Extract \( E \)

The extract function takes in a source \( D \) that is responsible for parsing the source of the conversational data (website, social media handle, wiki, forums etc.), perform data cleaning and preprocessing, and return \( E(D) \) which can be parsed by downstream NLP methods. In this paper, the purpose of our extract function is the parsing and cleansing of posts and comments from Nextdoor of several communities. Our extract function takes in a search keyword as input, then scrapes all available posts and respective comments with said keyword in the last one year to build a set of conversations. These conversations are then cleaned and tokenized. Furthermore, for use by downstream NLP models, we format each row of our dataset as follows:

\(<\text{SEP}>\text{keyword}</\text{SEP}>\)
\(<\text{SEP}>\text{post containing the keyword}</\text{SEP}>\)
\(<\text{SEP}>\text{reply to the post}</\text{SEP}>\)
3.2 Transform $T$

Let the result of the extract function be $E(D)$. For each row $T_i \in E(D)$, $T(T_i)$ transforms the representation of each post and reply pair embeddings into knowledge graph triples $(e_1, r, e_2)$ where $e_1, e_2$ are entities and $r$ is a relation defining those entities.

The transformation function is a sequence of two NLP open information extraction methods - (1) named entity recognition (NER), and (2) relation extraction (RE). For NER, we use a pre-trained RoBerta (Liu et al., 2019) language model fine tuned on the CoNLL-2003 dataset to perform named entity recognition. In order to understand the relation between the entities in the conversation using RE, we trained a MNLI-BART (Yin et al., 2019) to perform an entailment and contradiction a.k.a zero-shot classification in extrapolating the relationship between entities. In this method, we have a premise $P$ that is defined as the embedded conversation we have included in Section 3.1. Once we have embedded the entire text in the given schema we perform a zero-shot classification with two sets of labels. The first set of label attempts to define the relationship between entities in terms of emotions - (1) sadness, (2) anger, (3) joy, (4) optimism. The second set defines the type of relationship that exists within the concept of the community asset based model to identify in what context the entities are appearing. The community asset model are (1) public, (2) commercial, (3) human, (4) environmental, (5) social, (6) cultural.

In both cases we utilize the hypothesis The relationship between $e_1$ and $e_2$ is that of label, and allow the MNLI model to calculate the class probabilities of the candidate relations we provided. After we generated our label we mapped the resulting knowledge graph triple as $(e_1, \text{label}, e_2)$.

4 Case Studies

We have picked two topics of interest within the community to understand and show the usefulness of a knowledge graph in real life settings. The two topics are (1) welfare, and (2) education. We picked each of the topic due to their relevance in benefiting community development. Welfare is an approach to alleviating some of the issues particularly homelessness, that have increased even more due to the COVID-19 pandemic (Zufferey, CZ; Rodriguez et al., 2022; Benavides and Nukpezah, 2020). Education is always an important topic in a community, as residents are always striving towards better schools and career (Howard and Johansson, 2014).

In our extraction phase we have used keywords to get raw data for each of these topics and gather conversational data in the neighborhood. The keywords we used to extract community data regarding homelessness and welfare were - (1) homeless, (2) homeless shelters, and (3) welfare. The keywords we used to extract community data regarding education were - (1) education, and (2) school.

Initially, we performed a sentiment analysis of discussions regarding these keywords in the community. The sentiment analysis was performed by
running conversational data through our pipeline, with candidate relations being (1) sadness, (2) joy, (3) optimism, and (4) anger. As can be seen in Figure 4, there is almost equal amount of sadness and joy in discussions regarding homelessness and welfare. A deeper dive into the knowledge graph can give social scientists and community residents an understanding of why there is such an equal proportion of positivity and negativity. There is also healthy amount of optimism regarding welfare. In comparison, education has a significantly stronger presence of sadness in discussions in our target community. Further study could uncover what the reason is for such an approach - possible reasons could be discussions about the declining state of education or perhaps discussions related to schools resuming post-pandemic.

A primary advantage of building our knowledge graph is also the ease of downstream tasks. For example, we can create translational embeddings of the knowledge graph to allow visualization of the community knowledge in a two-dimensional space. As can be seen in Figure 5, we trained a TransE model (Bordes et al., 2013) to create embeddings for our knowledge graph. The embedding was then deployed in a Tensorboard Embedding Projector1. If we search for the nearest neighbors of schools, we get the names of several people in the educational board and also two schools that are rated good in the community as seen in Figure 5a. Nearest neighbors of homeless shelter gives the location of the shelter in the community along with the name of the person in charge of such services and welfare, as seen in Figure 5b. We did not curate what conversations we are capturing and neither did we include any demographic or locational data before building the knowledge graph through our pipeline, and our knowledge graph was able to display such contextual knowledge about the community by reading resident conversations.

The usefulness of knowledge graph does not stop at providing such visualizations and neighbor analysis. We can further synthesize knowledge graphs into community recommendation systems, community question answering systems, community assistants etc. to directly benefit data access and data literacy of the community residents. Our community knowledge graph lays the foundation for future applications to be built on top of for community learning and community knowledge sharing.

1https://projector.tensorflow.org/

Figure 4: Emotions expressed regarding the topics of (1) homelessness & welfare, and (2) education

5 Related Work

Our approach to defining the structure of our knowledge graph is centered on the community capital model paradigm to building sustainable and resilient communities (Callaghan and Colton, 2008). Existing literature studies this paradigm, emphasizing a data-driven approach in a multitude of settings both rural (Roberts and Townsend, 2016) and urban (Ivey and Bereitschaft, 2021). Particularly of note has been the increasing importance of improving social cohesion and capital i.e. encouraging participatory action research, improving data literacy, and improving social cohesion (Balestrini et al., 2017; Lalot et al., 2021; Chowdhury and Sharma, 2021). Our work focuses particularly on this area of reducing the technological barrier of allowing data-driven decision-making and better communication.

Recent lines of work have explored the knowledge graph completion tasks including entity extraction, linking, relation extraction, and knowledge graph population using NLP tasks. These have been applied to medical misinformation detection, fraudulent transaction detection. In most cases, such NLP tasks have been only applied to developing and furthering language models and language learning (Mitchell et al., 2018). Particularly of note are Stanford CoreNLP (Manning et al., 2014; Angeli et al., 2015) and entity pair classification using LUKE (Yamada et al., 2020). We subsequently adapt these core methodologies for our knowledge graph fact generation technique that requires minimal supervision and matches our ontology specific to community development.

Furthermore, in pushing the frontier of AI for positive impact, our community knowledge graph construction is related to the vast literature on computational modeling of existing problem domains...
(Chase et al., 2019; Bhatnagar et al., 2019; Booth et al., 2019; Abedin et al., 2019). Specifically focused on knowledge graph construction for modeling environmental, ecological, and urban networks (Alonso et al., 2019; Soares et al., 2019; Stewart and Liu, 2020; Ji et al., 2020; Cui et al., 2020; Lim et al., 2021) with a key difference that our technique will help at the hyper-local level of abstraction such as neighborhoods and communities within an urban environment.

6 Conclusion

In this paper, we introduce the task of community learning and establish a baseline methodology in approaching it. We discuss the importance of community development and its importance, and we show how conversations in social media contain contextual information regarding the community. We introduce the community knowledge graph with a schema based on top of community capital paradigm established in community development literature, and develop a pipeline to build such a knowledge graph. The pipeline we introduce in this paper is a series of extract, transform, and load tasks to parse conversational data and map it into our pre-defined community knowledge graph schema. We then perform qualitative analysis on two topics of interest to show what information can be gleaned from such a knowledge graph and future possibilities of this model. In the future, our community knowledge graph will be deployable and usable for public consumption in understanding communities and benefit them.

Limitations

In this paper we propose a schema for community knowledge graph, based on the community asset model in (Callaghan and Colton, 2008). However, that is a starting point for the development of a scalable and robust knowledge graph. Further work needs to go into defining a formal ontology that defines community (in the OWL 2.0 format). Ontologies such as KM4City and SCO 2.0 will definitely increase the interoperability, scalability and maintainability of our community knowledge graphs. We use a series of triples as knowledge graphs for the purposes of this experiment, in order to deploy our knowledge graph for real world applications it needs to have the ontology and a possible SPARQL endpoint for further queries and downstream tasks. The deployment of such a cyberinfrastructure is future work for the scope of this paper.

Theoretically, similar methods can be adapted to different domains and not community knowledge graphs. We discuss in Section 5 certain methods that are prevalent in literature. However, the applicability of our relation extraction approach in both community development and other domains needs to be studied further. happen.

While our paper performs only qualitative analysis of the community knowledge graph, further work in actual application systems of the knowledge graph will allow us to collect empirical data. This will provide us with a much richer quantitative analysis of the knowledge graph itself, which is lacking in this paper.
Ethical Considerations

The dataset we parsed initially contained user information which we have anonymized. To preserve the anonymity of the authors, we have also hidden location and community names from the paper and data discussions. Those will be released upon the completion of the blind-review process.

Due to the nature of the conversations happening in Nextdoor, despite our best efforts, our knowledge graph does store names of community residents that have been mentioned in conversation. This is unavoidable at this point in time and may have privacy concerns. That is one of the reasons we have decided against releasing the dataset publicly, but instead will provide the pipeline to create said dataset so that people only have access to Nextdoor conversations in their own communities.

The language model we use is built on top of pre-trained models RoBerta and BART, both of which have been trained on large corpus that have their own inherent biases. Our model as a result is subject to such bias. Furthermore, while we do use our analysis for positive applications and impacts these models can be utilized for harmful applications. As stated in (Cathy O’Neil, 2016), such big data approaches can cause more harm than good and has to be evaluated carefully.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Award No. 1943002

References

Alireza Abedin, S. Hamid Rezatofighi, Qinfeng Shi, and Damith C. Ranasinghe. 2019. SparseSense: Human Activity Recognition from Highly Sparse Sensor Data-streams Using Set-based Neural Networks. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, pages 5780–5786. International Joint Conferences on Artificial Intelligence Organization.

Omar Alonso, Vasileios Kandylas, and Serge-Eric Tremblay. 2019. Scalable Knowledge Graph Construction from Twitter. arXiv preprint arXiv:1906.05986.

Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D. Manning. 2015. Leveraging Linguistic Structure For Open Domain Information Extraction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 344–354. Beijing, China. Association for Computational Linguistics.

Mara Balestrini, Yvonne Rogers, Carolyn Hassan, Javi Creus, Martha King, and Paul Marshall. 2017. A city in common: a framework to orchestrate large-scale citizen engagement around urban issues. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pages 2282–2294.

Abraham David Benavides and Julius A. Nukpezah. 2020. How Local Governments Are Caring for the Homeless During the COVID-19 Pandemic. The American Review of Public Administration, 50(6-7):650–657.

Saumya Bhatnagar, Akshat Kumar, and Hoong Chuin Lau. 2019. Decision Making for Improving Maritime Traffic Safety Using Constraint Programming. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, pages 5794–5800. International Joint Conferences on Artificial Intelligence Organization.

Serena Booth, Christian Muise, and Julie Shah. 2019. Evaluating the Interpretability of the Knowledge Compilation Map: Communicating Logical Statements Effectively. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, pages 5801–5807. International Joint Conferences on Artificial Intelligence Organization.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In Advances in Neural Information Processing Systems, volume 26. Curran Associates, Inc.

Edith G Callaghan and John Colton. 2008. Building sustainable & resilient communities: a balancing of community capital. Environment, development and sustainability, 10(6):931–942. Publisher: Springer.

Cathy O’Neil. 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown, New York.

Jonathan Chase, Duc Thien Nguyen, Haiyang Sun, and Hoong Chuin Lau. 2019. Improving Law Enforcement Daily Deployment Through Machine Learning-Informed Optimization under Uncertainty. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, pages 5815–5821. International Joint Conferences on Artificial Intelligence Organization.

Md Towhidul A Chowdhury and Naveen Sharma. 2022. LotRec: A Recommender for Urban Vacant Lot Conversion. Publisher: arXiv Version Number: 1.

Md Towhidul Absar Chowdhury and Naveen Sharma. 2021. Citizenly: A platform to encourage data-driven decision making for the community by the
community. 2021 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybernetics (Cybernetics), pages 359–364. Publisher: IEEE.

Limeng Cui, Haeseung Seo, Maryam Tabar, Fenglong Ma, Suhang Wang, and Dongwon Lee. 2020. De- terrent: Knowledge guided graph attention network for detecting healthcare misinformation. Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pages 492–502.

UN Desa and others. 2016. Transforming our world: The 2030 agenda for sustainable development.

Linda Plitt Donaldson and Laura Daughtery. 2011. Introducing asset-based models of social justice into service learning: A social work approach. Journal of Community Practice, 19(1):80–99. Publisher: Taylor & Francis.

Sujan Dutta, Beibei Li, Daniel S. Nagin, and Ashiqur R. KhudaBukhsh. 2022. A Murder and Protests, the Capitol Riot, and the Chauvin Trial: Estimating Disparate Media Stance. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, pages 5059–5065, Vienna, Austria. International Joint Conferences on Artificial Intelligence Organization.

Emanuele Falzone, Riccardo Tommasini, Emanuele Della Valle, Petra Selmer, Stefan Plantikow, Hannes Voigt, Keith Hare, Ljubica Lazarevic, and Tobias Lindaaker. 2021. Semantic Foundations of Seraph Continuous Graph Query Language. arXiv preprint arXiv:2111.09228.

Howard Frumkin. 2014. Environmental metrics for community health. Community-Oriented Health Services: Practices Across Disciplines, page 135. Publisher: Springer Publishing Company.

Stephen Goldsmith and Susan Crawford. 2014. The Responsive City: Engaging Communities Through Data-Smart Governance. John Wiley & Sons. Google-Books-ID: ttRLBAAQBAJ.

Jenny Hong, Catalin Voss, and Christopher Manning. 2021. Challenges for Information Extraction from Dialogue in Criminal Law. In Proceedings of the 1st Workshop on NLP for Positive Impact, pages 71–81, Online. Association for Computational Linguistics.

Ann Howard and Magnus Johansson. 2014. Community Based Learning in Sweden and United States: what Works in Different Local Context? In Living Knowledge Conference, Copenhagen, Denmark (2014), pages 110–132. Aalborg University Copenhagen, Center for Design, Innovation and Sustainable . . .

Rebecca Ivey and Bradley Bereitschaft. 2021. The Impact of Walkability on the Sales Price of Commercial Properties When Controlling for the Effects of Economic Recession: A Case Study of Omaha, Nebraska. Journal of Real Estate Literature, 29(1):43–59. Publisher: Taylor & Francis.

Shaoxiong Ji, Shuirui Pan, Erik Cambria, Pekka Martinen, and Philip S Yu. 2020. A survey on knowledge graphs: Representation, acquisition and applications. arXiv preprint arXiv:2002.00388.

Zhijing Jin, Zeyu Peng, Tejas Vaidhya, Bernhard Schoelkopf, and Rada Mihalcea. 2021. Mining the Cause of Political Decision-Making from Social Media: A Case Study of COVID-19 Policies across the US States. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 288–301, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Rahim Kurwa. 2019. Building the Digitally Gated Community: The Case of Nextdoor. Surveillance & Society, 17(1/2):111–117.

Fanny Lalot, Dominic Abrams, Jo Broadwood, Kaya Davies Hayon, and Isobel Platts-Dunn. 2021. The social cohesion investment: Communities that invested in integration programmes are showing greater social cohesion in the midst of the COVID-19 pandemic. Journal of Community & Applied Social Psychology. Publisher: Wiley Online Library.

Katie Lambright. 2019. Digital Redlining: The Nextdoor App and the Neighborhood of Make-Believe. Cultural Critique, 103(1):84–90. Publisher: University of Minnesota Press.

Mei Qi Lim, Xiaonian Wang, Oliver Inderwildi, and Markus Kraft. 2021. The World Avatar—a world model for facilitating interoperability.

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. Publisher: arXiv Version Number: 1.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pages 55–60, Baltimore, Maryland. Association for Computational Linguistics.

Christina A. Masden, Catherine Grevet, Rebecca E. Grinter, Eric Gilbert, and W. Keith Edwards. 2014. Tensions in scaling-up community social media: a multi-neighborhood study of nextdoor. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 3239–3248, Toronto Ontario Canada. ACM.
Tom Mitchell, William Cohen, Estevam Hruschka, Partha Talukdar, Bishan Yang, Justin Betteridge, Andrew Carlson, Bhavana Dalvi, Matt Gardner, Bryan Kisiel, and others. 2018. Never-ending learning. Communications of the ACM, 61(5):103–115. Publisher: ACM New York, NY, USA.

Shriphani Palakodety, Ashiqur R. KhudaBukhsh, and Jaime G. Carbonell. 2019. Hope Speech Detection: A Computational Analysis of the Voice of Peace. Publisher: arXiv Version Number: 4.

Elisabeth Roberts and Leanne Townsend. 2016. The contribution of the creative economy to the resilience of rural communities: exploring cultural and digital capital. Sociologia Ruralis, 56(2):197–219. Publisher: Wiley Online Library.

Natalia M. Rodriguez, Rebecca G. Martinez, Rebecca Ziolkowski, Celia Tolliver, Hope Young, and Yumary Ruiz. 2022. “COVID knocked me straight into the dirt”: perspectives from people experiencing homelessness on the impacts of the COVID-19 pandemic. BMC Public Health, 22(1):1327.

Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the Blanks: Distributional Similarity for Relation Learning. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 2895–2905.

Michael Stewart and Wei Liu. 2020. Seq2KG: An End-to-End Neural Model for Domain Agnostic Knowledge Graph (not Text Graph) Construction from Text. In Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning, volume 17, pages 748–757. Issue: 1.

Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. Journal of machine learning research, 9(11).

Ricardo Vinuesa, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark, and Francesco Fuso Nerini. 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications, 11(1):233. Number: 1 Publisher: Nature Publishing Group.

Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. 2020. Luke: deep contextualized entity representations with entity-aware self-attention. arXiv preprint arXiv:2010.01057.

Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. arXiv preprint arXiv:1909.00161.

Clay H. Yoo, Shriphani Palakodety, Rupak Sarkar, and Ashiqur KhudaBukhsh. 2021. Empathy and Hope: Resource Transfer to Model Inter-country Social Media Dynamics. In Proceedings of the 1st Workshop on NLP for Positive Impact, pages 125–134, Online. Association for Computational Linguistics.

Chloé Zufferey (CZ), Morten Skovdal (MS), Iben Mundbjerg Gjødsbøl (IMG), and Signe Smith Jervelund (SSJ). 2022. Caring for people experiencing homelessness in times of crisis: Realities of essential service providers during the COVID-19 pandemic in Copenhagen, Denmark. International Journal of Disaster Risk Reduction, 79:103157.