

LEVERAGING SOCIAL MEDIA AND NATURAL LANGUAGE PROCESSING FOR UNDERSTANDING CULTURAL PERCEPTION IN URBAN RENEWAL

Insights from Nantou Walled City

SHUIYANG LIN¹ and JEROEN VAN AMEIJDE²

^{1,2}*The Chinese University of Hong Kong.*

¹*lynnsylin@link.cuhk.edu.hk, 0000-0002-3473-688X*

²*jeroen.vanameijde@cuhk.edu.hk, 0000-0002-3635-3305*

Abstract. The preservation of cultural heritage during urban renewal processes is an issue that is often subservient to economic urban development goals. While physical structures of historic importance might be preserved or adapted, the social and cultural significance of historic urban places is often lost during the comprehensive redevelopment of historic neighbourhoods. The absence of adequate measurements and procedures to assess the cultural and social values of historic districts remains a key concern in urban renewal practice. Recent advances in Big Data analysis of social media posts through Machine Learning-based Natural Language Processing (NLP) tools offer ways to quantify the social and cultural perception of historic urban districts and guide culturally sensitive strategies for revitalization. This research presents a novel workflow to analyse text data from social media platforms. Referencing recently developed methods for multidimensional text mining and sentiment analysis, it obtains online text data by identifying the keywords on the Weibo platform and applies BERTopic to construct a quantitative assessment of cultural perceptions by identifying the topics and sentiments about a historic district. Sentiment analysis tools were used to measure emotional polarity in posts to understand social media users' subjective views towards various urban places. Using the case study of Nantou Walled City in Shenzhen, a comparative examination of social media engagement before and after district renewal sheds light on the interplay between social media participation and historic district characteristics. The method presented in this study enables future studies and practices in urban planning and cultural heritage conservation which adopt a comprehensive consideration of resident and community perspectives in urban renewal, to achieve more holistic and sustainable urban development.

Keywords. Natural language processing, social media data analysis, cultural heritage, sentiment analysis, BERTopic.

1. Introduction

Within the context of global urban renewal, there has been a paradigm shift from a singular focus on economic development to a greater emphasis on cultural preservation and diversity in communities (Gittell, 1992; Bailey, Miles & Stark, 2004; Hsing, 2010). Particularly in China, with a decreasing scope for urban expansion and the direction of urban regeneration shifting towards a 'quality-driven' approach, the rational protection of cultural heritage and land redevelopment has become the focus of attention (Wang, 2009; Zhong, 2016; Ding & Wu, 2017). However, how to find a balance between rapid urban development and cultural heritage protection remains a complex challenge.

Currently, traditional methods for assessing cultural values, such as the study of historical documents, expert interviews, and on-site research, are often insufficient to meet the challenge because they are usually labour-intensive and time-consuming (Harry & Klingner, 2006; Ujang & Zakariya, 2015). With the development of the field of artificial intelligence, Natural Language Processing (NLP) techniques based on deep learning provide new perspectives and tools for urban design, especially in analysing large-scale textual data and capturing public perceptions (Batrinca & Treleven, 2015; Jang & Kim, 2019). Applying this technology to the field of urban design means that a large amount of under-utilised textual data (e.g., planning documents, resident feedback, or social media content) in urban design can be effectively utilised, providing new opportunities for urban researchers to quantitatively analyse the public's cultural perceptions of historic urban areas (Fuchs, 2014; Hirschberg & Manning, 2015; Huai et al., 2023).

As an important historical and cultural heritage site in Shenzhen, Nantou Walled City's cultural preservation and renewal as part of Shenzhen's wider urban development has triggered intense discussions (Zordan et al., 2019). Since its hosting of the Bi-City Biennale of Urbanism/Architecture (UABB) in 2017, it provides an important opportunity to explore how urban micro-regeneration can be achieved through cultural interventions. In response to this opportunity, this study has used NLP-based methods to analyse public discussions on Weibo and conducted a multi-dimensional thematic quantitative analysis of cultural perceptions of Nantou Walled City. Through in-depth analysis of social media text data, the study reveals the public's cognitive and emotional responses to the cultural values of Nantou Walled City and tracks the trends of these perceptions after the important cultural events of the Biennale. Based on these methods, this paper presents a novel assessment process, which provides a new perspective for urban planners and conservationists in the assessment of cultural values and the development of intervention strategies in historic urban areas.

2. Methods

2.1. EVALUATION FRAMEWORK

To structure the assessment of cultural perception of historic districts, this study has conducted a review of the relevant topics related to cultural perception, as well as methods for the evaluation of emotional polarity and the change in social media participation. A framework and process for evaluating the cultural perception of historic districts based on social media data was constructed and is shown in Figure 1.

The framework contains five steps, namely [1] comment data crawling. [2] Text Cleaning and Jieba Segmentation. [3] Topic modelling using BERTopic to obtain topics and keywords related to cultural perception. [4] Bert-based fine-grained sentiment analysis construction. [5] Sentiment polarity and social media engagement evolution analysis over time.

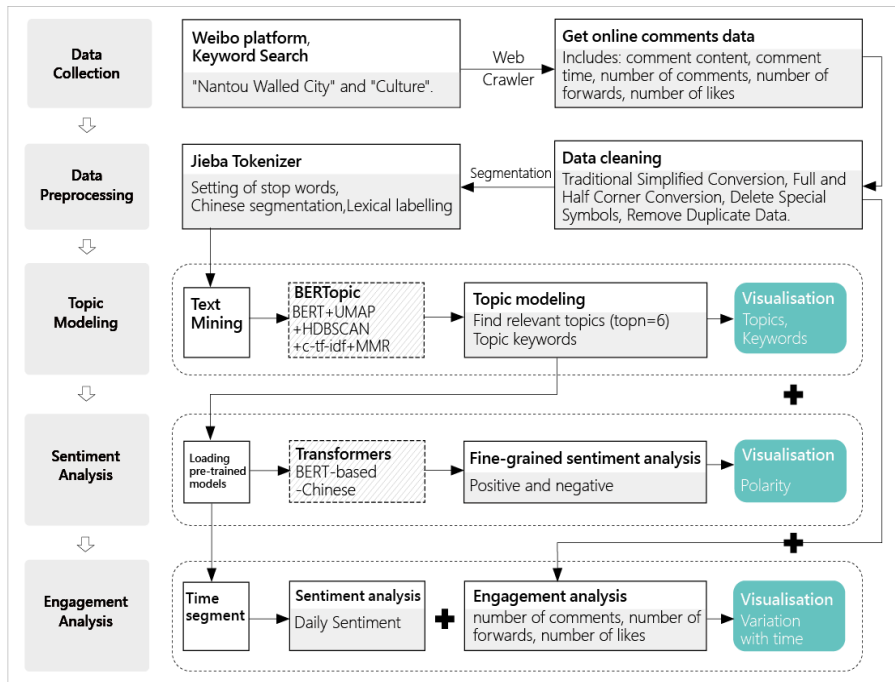


Figure 1. An evaluation framework for cultural perception based on social media data

In this study, we obtained the user interaction data of Nantou Walled City within the Chinese region from the Weibo platform as the main source of research data from 1 January 2023 to 1 December 2023. By setting the keywords "Nantou Walled City" and "culture" on the Weibo platform, a total of 5,702 comments were collected using web crawler technology, involving the online comment content, comment time, retweet volume, comment volume, and likes volume.

In the data pre-processing stage, screening measures were taken to exclude comments with obvious advertising colours, comments that only mentioned Nantou Walled City, but the actual description of the content or object of the evaluation was not related to culture and duplicated content, and finally, 3,037 valid text data were screened out. Text cleaning includes traditional and simplified conversion and blank character removal. The Jieba tokenizer was used for lexical annotation, and combined with the stop word list provided by the Sichuan University Machine Intelligence Laboratory, the Harbin Institute of Technology, and Baidu, auxiliary words, conjunctions, and punctuation marks were removed to ensure the quality of data

cleaning.

2.2. TOPIC MODELLING AND FINE-GRAINED SENTIMENT ANALYSIS

After completing the initial data screening and pre-processing, this study used the BERTopic model to analyse the topic modelling of social media texts related to Nantou Walled City to reveal the public's cultural perceptions. By integrating the advanced language understanding capabilities of deep learning with topic modelling techniques, BERTopic is able to discover implicit topics in textual data without the need for pre-defined labels (Abuzayed & Al-Khalifa, 2021; Guan et al., 2022; Liu et al., 2016), which is particularly suitable for social media texts, effectively capturing subtle topic differences and revealing semantic and contextual depth (Egger & Yu, 2022; Grootendorst, 2022). Compared to traditional LDA topic models, NMF data feature extraction, and SVM classification algorithms, BERTopic utilises BERT deep learning text embedding to provide a more granular analytical perspective for understanding complex user opinions and sentiments on social media (Reisenbichler & Reutterer, 2018; Subakti et al., 2022; Young et al., 2018).

The model combines the techniques of BERT text embedding, UMAP dimensionality reduction, and HDBSCAN clustering to effectively extract text topics and their keywords. Further, the keywords were refined by the c-TF-IDF algorithm and MMR algorithm to ensure the quality of the keywords. The analysis reveals the topics that are closely related to cultural perception, and the `find_topics` function is used to find the six most relevant themes to the cultural perception theme. Finally, a comprehensive dataset including text content, topic number, keywords, and similarity was formed. In processing about 3037 valid comment data related to the Nantou Walled City collected from microblogging platforms, the C-TF-IDF algorithm in BERTopic is able to produce dense topic clustering that effectively captures subtle topic differences. This is especially crucial for revealing the complex emotions and perceptions of the public towards cultural heritage. In contrast, traditional LDA models may not be sensitive enough to handle the context and complexity of language, while SVM, as a classification algorithm, has limited applicability in topic discovery tasks with unlabelled data.

Following the completion of BERTopic-based topic modelling, this study further develops fine-grained sentiment analysis using the BERT-based Chinese model, aiming to accurately explore and reveal the public's sentiment feedback on each topic of the ancient city of Nantou. The text input is optimised by BERT, and the positive and negative sentiments are quantitatively assessed by gradient-free computation and softmax function to ensure accurate prediction and quantitative quantification of textual sentiment tendencies.

2.3. DAILY SENTIMENT AND ENGAGEMENT ANALYSIS

In this step, temporal data was introduced to calculate sentiment scores and engagement daily. Specifically, daily sentiment polarity is derived by calculating the difference between the average of the positive and negative sentiment scores of all commented content on that day, thus quantifying the overall trend of public sentiment on that day. Meanwhile, daily engagement was measured by aggregating the number

of retweets, comments, likes, and comment likes for all posts on that day to reflect the frequency of public interaction on social media on relevant cultural topics.

3. Analysis and Results

3.1. TOPIC MODELLING AND FINE-GRAINED SENTIMENT ANALYSIS

This study provides an in-depth analysis of public attitudes towards the perception of cultural heritage through BERTopic topic modelling of social media texts related to the Nantou Walled City. By analysing the microblog comments, this study identifies themes that are closely related to cultural perceptions and explores in detail the keywords and their distribution probabilities for each topic.

As shown in Figure 2, the visualised distribution of the topics after dimensionality reduction clearly shows the aggregation of each topic in the low-dimensional space, as well as the keywords of each topic and their corresponding probability values, which reflect the interrelationships between different cultural perceptions topics and their prominence in the public discussion, as well as the weights of the different keywords in each topic. A total of 16 topics related to cultural perception were generated, of which only the six most relevant topics and the topic keywords were displayed according to the similarity ranking (Table 1). Topic 9 (History and Architecture) appeared with relatively high similarity, and its keywords "history, culture, ancient city, utility, architecture" highlighted the public's concern and discussion about the historical value and architecture of Nantou Walled City. Topic 16 (Art and Creative Products) reveals a trend of cultural exchange related to cultural and creative products, with keywords such as "creativity, design, cultural and creative products, handmade" highlighting contemporary consumers' preference for goods with cultural meanings. In addition, Topic 12 (Local specialties of food) and Topic 15 (Art Installations and Exhibitions) reflect public participation in cultural activities, thus demonstrating the integration and influence of cultural activities in urban life.

From the distribution of topics and keywords, we can conclude that the public discussion of cultural perceptions focuses on historical values and architecture (Topic 9), art and creative products (Topic 16), local specialties of food (Topic 12), art installations and exhibitions (Topic 15), youth culture and music activities (Topic 10), and tourism perceptions and experiences (Topic 5). These findings suggest that cultural perceptions on social media show a common focus on traditional and modern elements, reflecting a high level of public interest in the preservation of cultural heritage and modern cultural creativity.

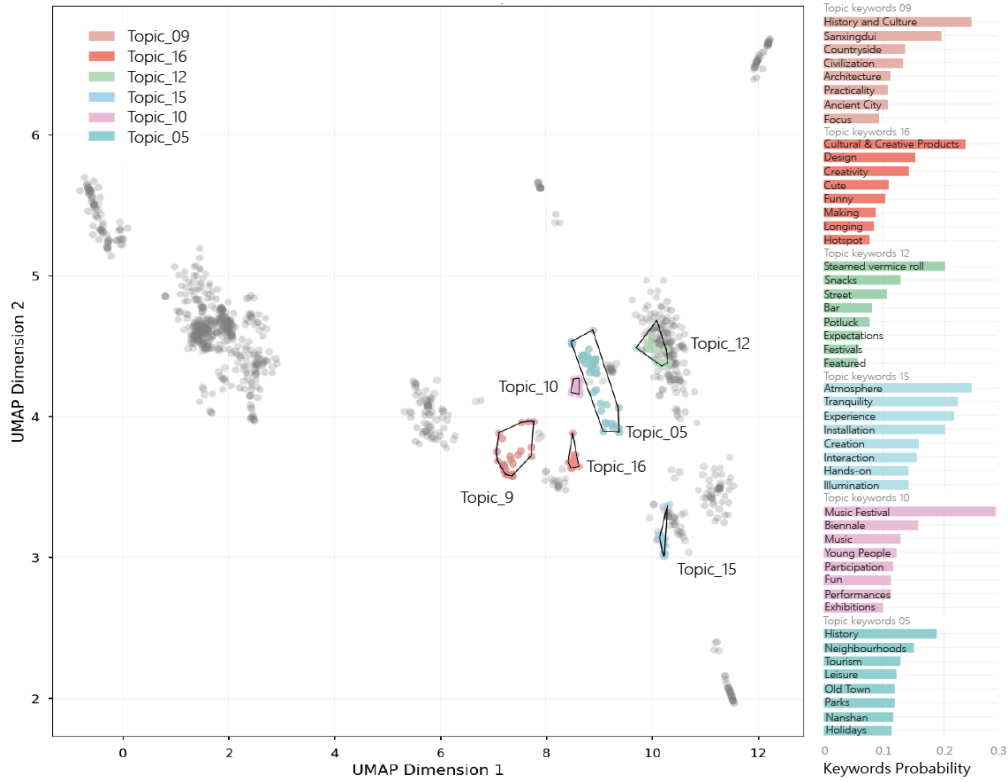


Figure 2. Distribution of topics related to cultural perception and topic keyword probabilities

Table 1. Topic similarity and keyword list

Topic	Similarity	Keywords
9	0.729352	History and Culture, Sanxingdui, Countryside, Civilisation, Architecture, Practicality, Ancient City, Focus, Culture, Places
16	0.566669	Cultural and Creative Products, Design, Creativity, Cute, Funny, Making, Longing, Hotspot, Bracelet, Handmade
12	0.461585	Steamed vermicelli roll, Snacks, Street, Bar, Potluck, Expectations, Festivals, Featured, Firecrackers, Moonwatching
15	0.458987	Atmosphere, Tranquility, Experience, Installation, Creation, Interaction, Hands-on, Illumination, Tumbling, Interaction
10	0.443497	Music Festivals, Biennale, Music, Young People, Participation, Fun, Performances, Exhibitions, Happy, Likes
5	0.440511	History, Neighbourhoods, Tourism, Leisure, Old Town, Parks, Nanshan, Holidays, Festivals, Surroundings

Further fine-grained sentiment analyses quantified the sentiment tendencies of the topics (Figure 3), revealing that for example, in Topic 9, although the "History and Architecture" topic was discussed with great enthusiasm, it was accompanied by significantly negative sentiments, reflecting concerns about cultural heritage management and preservation measures. Sentiment scores for Topic 16, Arts and Creative Products, on the other hand, were relatively balanced but tended to be slightly negative, possibly revealing misgivings or dissatisfaction with certain aspects of the creative industries and local cultural offerings. Overall, these results show complex emotional feedback from the public on cultural phenomena, with both positive resonance and critical voices, revealing comprehensive emotional feedback from the public on cultural phenomena.

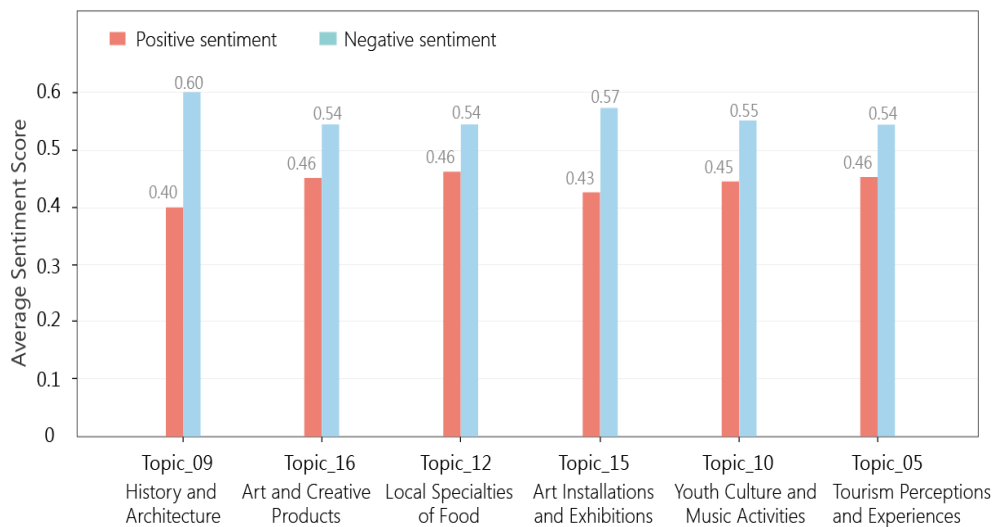


Figure 3. Fine-grained sentiment analysis

3.2. SENTIMENT AND ENGAGEMENT UNDER TIME SERIES

Through time-series analyses, this study deepens the insights into the emotional polarity and public engagement on the topic of cultural perception of the Nantou Walled City. As shown in Figure 4, before the launch of the Architecture Biennale, the affective polarity index was relatively stable, reflecting little fluctuation in public sentiment towards the topic of cultural perception. However, with the launch of the Biennale, the affective polarity index fluctuated, revealing that different exhibitions and events inspired diverse emotional responses. In particular, the elevated negative sentiment scores during some periods (27 October to 15 November) hint at critical public feedback on certain exhibitions or events. This may be related to content related to history and cultural heritage (Topic 9) and art installations and exhibitions (Topic 15) that did not meet the public's expectations or elicited negative comments.

In terms of engagement, spikes during the Biennale reflect the fact that certain major cultural events triggered extensive interaction on social media. These high points

of engagement not only signalled the success of public participation in cultural events but also highlighted the impact of activities such as openings, special exhibitions, or artist interactions among the public.



Figure 4. Evolution of sentiment polarity and engagement over time

4. Discussion and Conclusion

This study quantitatively assesses the perceived cultural value of Nantou Walled City by using advanced natural language processing techniques and social media analytics tools. Within the context of finding a balance between economic development and cultural heritage preservation, this study found that during cultural events such as the Biennale, the public's emotional polarity, and engagement showed significant fluctuations, reflecting the public's sensitive reaction to the content of the event. Particularly during the thematic discussions on history and cultural heritage, negative emotions dominated, possibly suggesting the public's scepticism towards cultural preservation strategies. Furthermore, although public participation in cultural events increased, the duality of affective feedback revealed complex public attitudes towards the content of cultural exhibitions and events.

Although this study provided valuable insights, the research methodology still needs to be improved in the face of challenges specific to Chinese social media data, such as information filtering and the accuracy of sentiment analysis. Future work should focus on improving the accuracy and automation of data processing, including refining culturally-aware lexicons and utilising big data techniques to improve the efficiency of model training. Beyond social media text data, the integration of other types of data sources, e.g. public policy documents, news articles, and so on, may provide a more comprehensive view of cultural perceptions. Furthermore, the introduction of sociological and anthropological perspectives and the use of qualitative research methods (such as interviews or case studies) combined with existing

quantitative analyses can provide a deeper understanding of the impact of cultural redevelopment. Through these improvements, researchers will be able to more accurately capture and understand the public's cultural perceptions of historic urban areas and provide more precise strategies for cultural heritage protection in urban renewal.

References

- Abuzayed, A. and Al-Khalifa, H. (2021). Bert for arabic topic modeling: an experimental study on bertopic technique. *Procedia Computer Science*, 189, 191-194. <https://doi.org/10.1016/j.procs.2021.05.096>
- Bailey, C., Miles, S., & Stark, P. (2004). Culture-led urban regeneration and the revitalisation of identities in Newcastle, Gateshead, and the North East of England. *International Journal of Cultural Policy*, 10(1), 47–65.
- Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: A survey of techniques, tools and platforms. *AI & Society*, 30(1), 89-116.
- Ding, F., & Wu, J. (2017). The evolution of the concept of urban regeneration and the practical connotation in today's society. *Urban Planning Forum*, 6,87–95.
- Egger, R. and Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in Sociology*, 7. <https://doi.org/10.3389/fsoc.2022.886498>
- Fuchs, C. (2014). Social Media and the Public Sphere. *TripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 12(1), 57–101. <https://doi.org/10.31269/triplec.v12i1.552>
- Gittell, R. J. (1992). *Renewing cities*. Princeton: Princeton University Press.
- Grootendorst, M. (2022). Bertopic: neural topic modeling with a class-based tf-idf procedure. <https://doi.org/10.48550/arxiv.2203.05794>
- Guan, R., Zhang, H., Liang, Y., Giunchiglia, F., Huang, L., & Feng, X. (2022). Deep feature-based text clustering and its explanation. *Ieee Transactions on Knowledge and Data Engineering*, 34(8), 3669-3680. <https://doi.org/10.1109/tkde.2020.3028943>
- Harry, B., & Klingner, J. (2006). The limits of survey research methods in assessing the problem of minority overrepresentation in special education. *Exceptional Children*, 72(2), 163-177.
- Hirschberg, J., Manning, C.D., 2015. Advances in natural language processing. *Science* 349 (6245), 261–266.
- Hsing, Y.-t. (2010). *The Great Urban Transformation: Politics of Land and Property in China*. Oxford University Press.
- Huai, S., Liu, S., Zheng, T., & Van de Voorde, T. (2023). Are social media data and survey data consistent in measuring park visitation, park satisfaction, and their influencing factors? A case study in Shanghai. *Urban Forestry & Urban Greening*, 81, 127869. <https://doi.org/10.1016/j.ufug.2023.127869>
- Jang, K.M., Kim, Y., (2019). Crowd-sourced cognitive mapping: a new way of displaying people's cognitive perception of urban space. *PloS One* 14 (6).
- Liu, L., Tang, L., Wen, D., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *Springerplus*, 5(1). <https://doi.org/10.1186/s40064-016-3252-8>
- Reisenbichler, M. and Reutterer, T. (2018). Topic modeling in marketing: recent advances and research opportunities. *Journal of Business Economics*, 89(3), 327-356. <https://doi.org/10.1007/s11573-018-0915-7>
- Subakti, A., Murfi, H., & Hariadi, N. (2022). The performance of bert as data representation of text clustering. *Journal of Big Data*, 9(1). <https://doi.org/10.1186/s40537-022-00564-9>

- Ujang, N., & Zakariya, K. (2015). The notion of place, place meaning, and identity in urban regeneration. *Procedia - Social and Behavioral Sciences*, 170, 709-717.
- Vallance, S., Perkins, H. C., & Dixon, J. E. (2011). Social sustainability, past and future: Undoing unintended consequences for the Earth's survival. *International Journal of Sustainable Development*, 14(1/2), 121-135.
- Wang, J. (2009). "Art in the capital": Shaping distinctiveness in a culture-led urban regeneration project in Red Town, Shanghai. *Cities*, 26(6), 318-330.
- Young, T., Hazarika, D., Poria, S., & Wang, Z. (2018). Recent trends in deep learning based natural language processing [review article]. *Ieee Computational Intelligence Magazine*, 13(3), 55-75. <https://doi.org/10.1109/mci.2018.2840738>
- Zhong, S. (2016). Artists and Shanghai's culture-led urban regeneration. *Cities*, 56, 165-171.
- Zordan, M., Talamini, G., & Villani, C. (2019). The Association between Ground Floor Features and Public Open Space Face-To-Face Interactions: Evidence from Nantou Village, Shenzhen. *International Journal of Environmental Research and Public Health*, 16(24), 4934. <https://doi.org/10.3390/ijerph16244934>