**Appendix 2: Validation of the LDA model**

To assess external validity, we firstly considered the timing of increases and declines in the prevalence of topics, as suggested by Di Maggio et al (2013). Figure 1 shows the monthly variations in the percentage of tweets corresponding to each topic. For example, the topics related to COVID-19 mainly (4, 9, 10 and 14), occupy more attention from January 2020, progressively declining as the pandemic stabilises in the following months. Topic 2 concerning Hong Kong increases in May, when the National Security Law was promulgated, and keeps saliency over June and July, coinciding with its approval and the Special Administrative Region Establishment Day. Held on July 1st, it commemorates the anniversary of the transfer of sovereignty over Hong Kong from the United Kingdom to China. The topic that highlights the economic recovery (16) has a growing evolution once the worst months of the pandemic were overcome. Finally, topic 5, dealing with celebrations and festivities, has a higher saliency in the months of January 2020 and February 2021, coinciding with the Chinese new years, and in October, when the mid-Autumn festival was celebrated. Thus, these examples show that the prevalence of certain topics is time-sensitive in predictable ways, corroborating the validity of our model.

**Figure 1: Prevalence of topics each month**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

This validation is further supported statistically using the Spearman correlation (Spearman, 1904). Figure 2 shows a coefficient where -1 indicates a total negative correlation and 1 indicates a complete positive correlation between the number of tweets of two topics with respect to time. When it is positive, it indicates that as one topic has grown, the other topic has also grown in the proportion indicated by the value of the correlation. When it is negative, it indicates that when one topic has been discussed more, the other topic has been discussed less. Through this coefficient we can verify the relation between the increases and the declines in the prevalence of topics through time. As expected, semantically connected topics have a strong positive correlation, as for example, topics mainly related to COVID-19 (4, 9, 10 and 14). On the contrary, Topic 16, associated with the economic recovery, has an expected negative correlation with the COVID-19 topics: the salience of any of them implies less focus on topic 16. Topic 16, however, is strongly correlated with topics about development and infrastructure (1, 3, 7), and cooperation and provision of vaccines (6). These five latter topics also have a strong mutual correlation, suggesting a link between Chinese economic prosperity, multilateralism and supply of infrastructure and health resources.

**Figure 2: Correlation among topics**

**