1. Using the page below, we can obtain 6 forums,

[1] <https://news.discuss.com.hk/forumdisplay.php?fid=1230>

[2] <https://news.discuss.com.hk/forumdisplay.php?fid=1136>

[3] <https://news.discuss.com.hk/forumdisplay.php?fid=54>

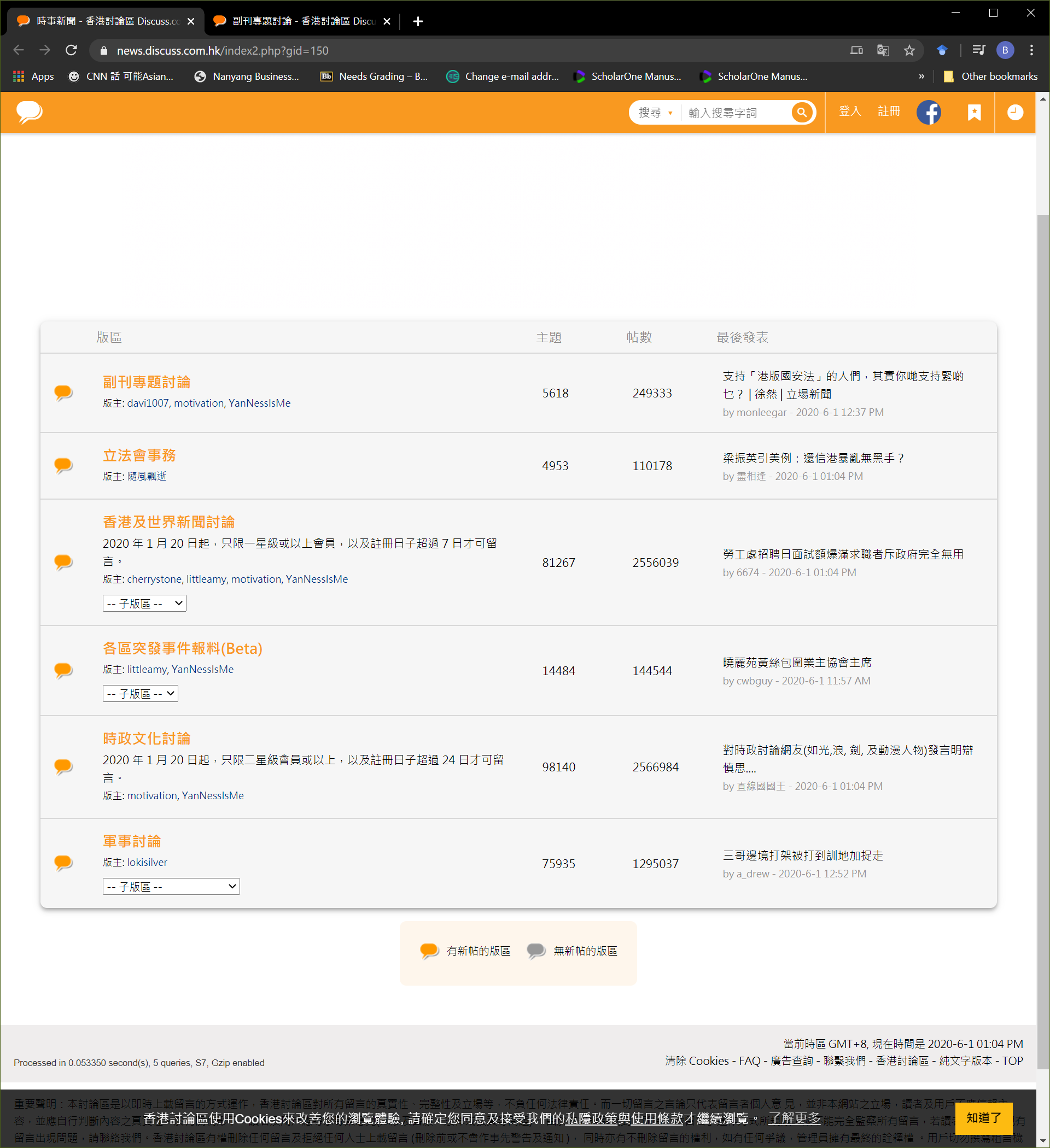
[4] <https://news.discuss.com.hk/forumdisplay.php?fid=1121>

[5] <https://news.discuss.com.hk/forumdisplay.php?fid=106>

[6] <https://news.discuss.com.hk/forumdisplay.php?fid=193>

To Capture: We need to capture posting within each of the 6 forums

<https://news.discuss.com.hk/index2.php?gid=150>



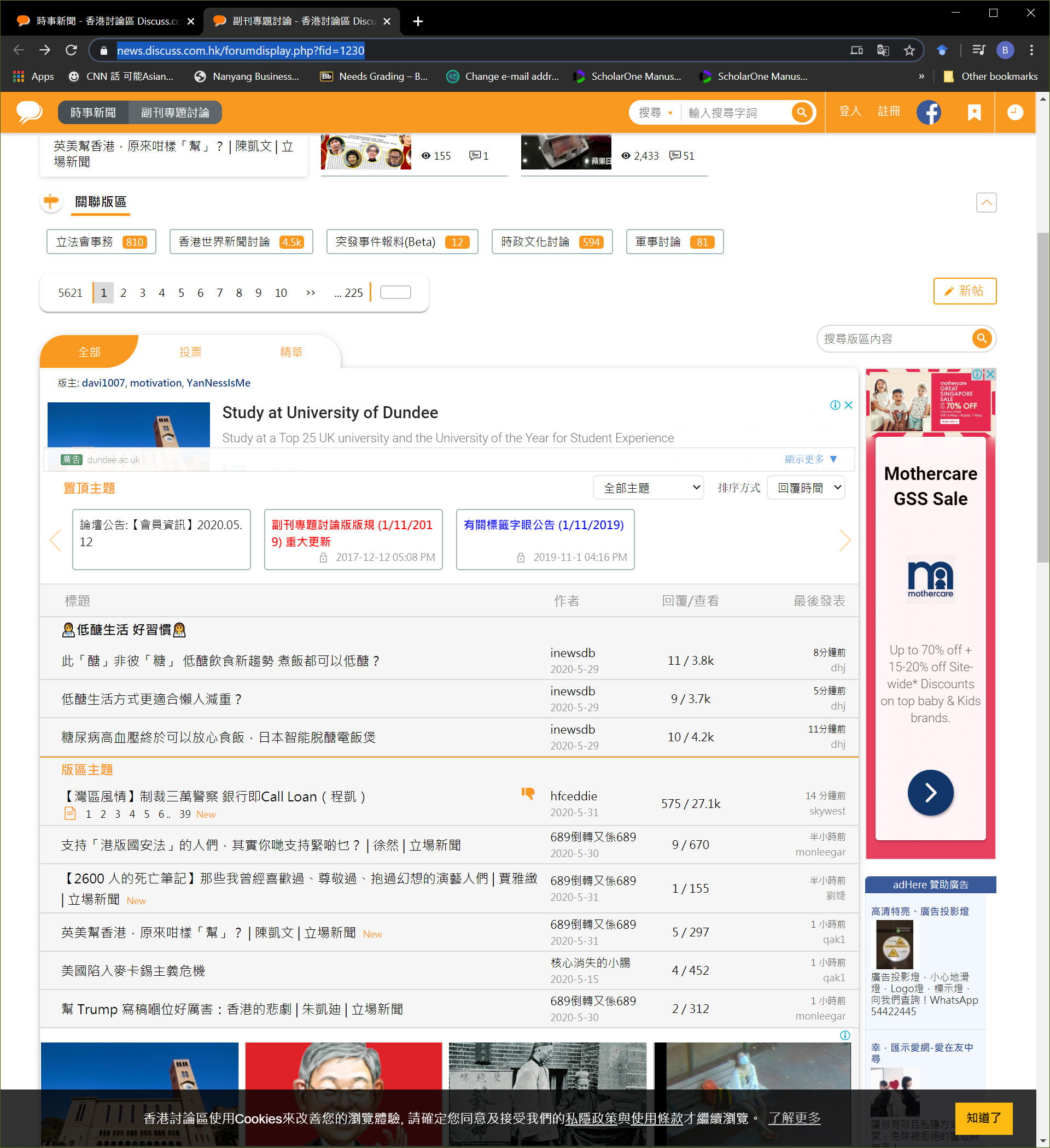
2. Using [1] as an example,

<https://news.discuss.com.hk/forumdisplay.php?fid=1230>

Each of the pages within [1] can be navigated using the following URL structure:

https://news.discuss.com.hk/forumdisplay.php?fid=1230&page=X

X can be replaced with 1… 225 (as in the example below); if >225 is used, page 225 will be displayed



A

B

1

2

A = list of pages under this forum

B = list of threats within this page

1 = Each of the threats, for example <https://www.discuss.com.hk/viewthread.php?tid=29108306&extra=page%3D1>

To capture

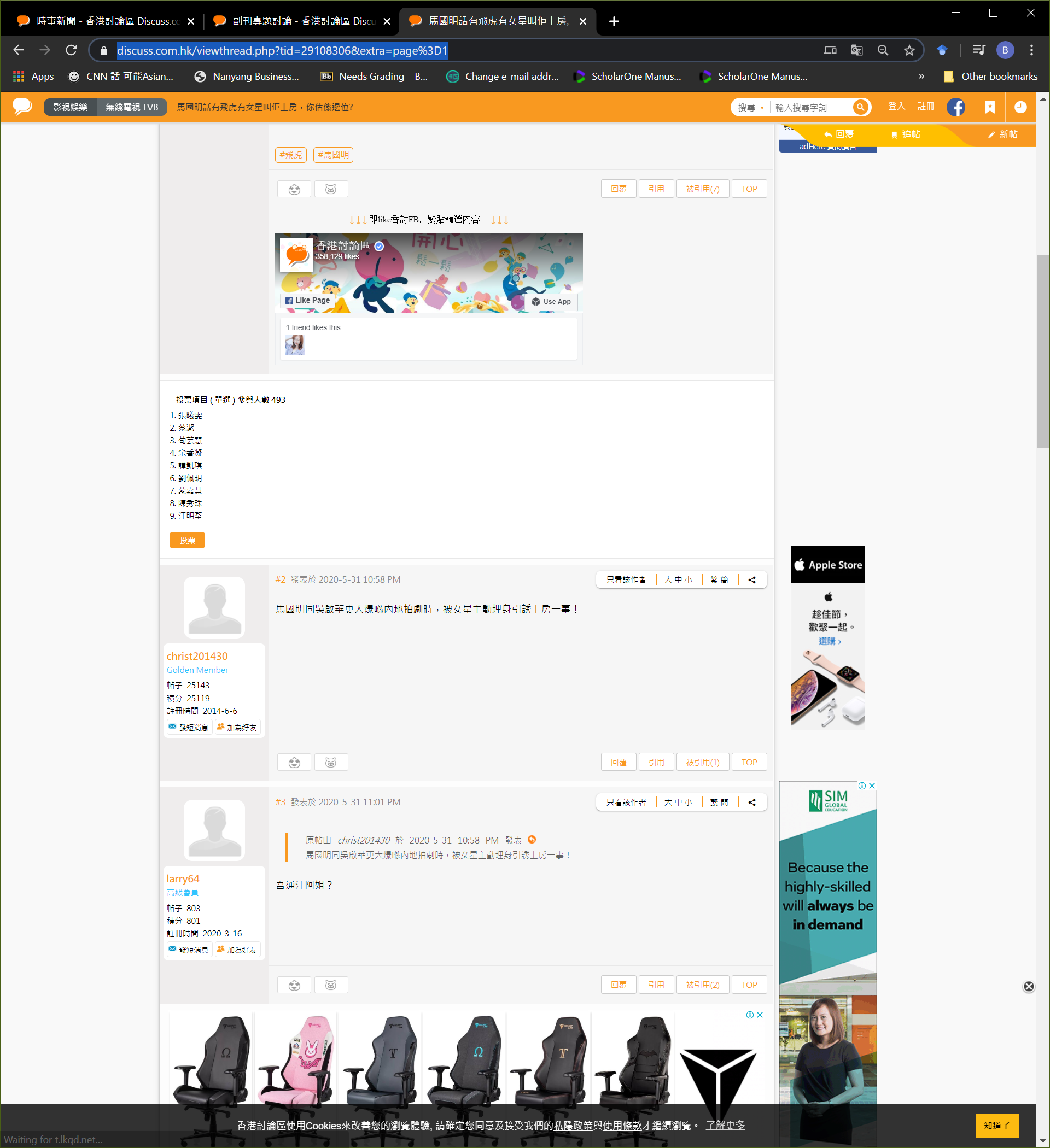
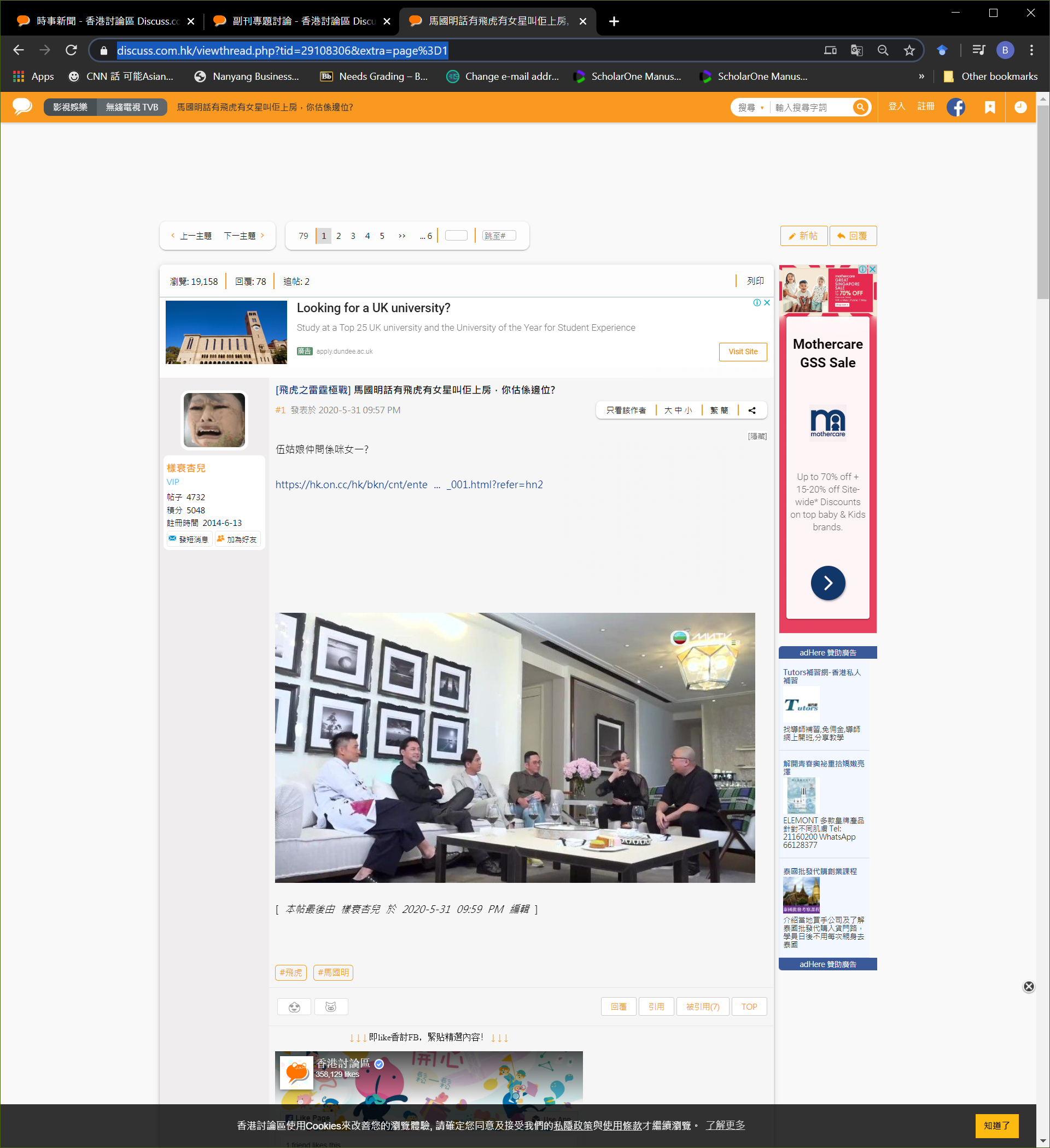
2: Responses / Views

3. For example, take <https://www.discuss.com.hk/viewthread.php?tid=29108306&extra=page%3D1>

Each of the pages within this threat can be navigated using the following URL structure:

[https://www.discuss.com.hk/viewthread.php?tid=29108306&extra=page%3D1&page=](https://www.discuss.com.hk/viewthread.php?tid=29108306&extra=page%3D1&page=2)X

X can be replaced with 1… 6 (as in the example below); if >6 is used, page 6 will be displayed



A

B

B3

B1

B2

B4

C

C1

C2

A = number of pages within this threat

To capture

B: The 1st post of a threat (this post might appear as the top post in all the pages of this threat)

B1 = Threat title

B2 = Sequence number of posting (#1) and posting date/time

B3 = author of the post, user type (VIP), amount of posting, amount of points, registration date

B4 = content of the post

C = the 2nd post of this threat

C1 = same as B3

C2 = same as B4