

**gml:** final gml files dating from 24 Sep to 30 Oct, generated from txt files in text folder

**text:** Twitter contents by date after cleaning data sets from jhy and qhd

**processing-codes:** python scripts used to clean data and generate gml files

**pickle:** files that can be read directly into Pandas DataFrame. Files are corresponding to the gml files, and the consolidated\_data.pkl is the data for the 34 networks combined together

#### **jhy data set timeline :**

ignore file 15 because too small

1. Sep 30 09:58:44 - Sep 30 20:38:56
2. Oct 01 08:17:38 - Oct 01 19:08:22
3. Oct 03 06:28:38 - Oct 04 03:34:03
4. Oct 04 03:50:48 - Oct 05 02:38:42
5. Oct 05 02:51:33 - Oct 05 03:04:21
6. Oct 06 02:52:36 - Oct 06 02:59:29
7. Oct 07 02:52:51 - Oct 07 03:38:33
8. Oct 08 02:53:09 - Oct 11 06:40:40
9. Oct 09 02:53:24 - Oct 11 06:40:46
10. Oct 10 03:22:01 - Oct 14 06:00:30
11. Oct 14 06:05:23 - Oct 15 15:08:36
12. Oct 16 08:34:53 - Oct 16 12:36:27
13. Oct 17 07:04:16 - Oct 20 08:35:08
14. Oct 20 08:37:20 - Oct 21 20:24:11
15. NA
16. Oct 24 05:36:55 - Oct 30 10:01:12

#### **qhd data set timeline:**

missing twitter data from 4, 12, 13,16, 17, 22 Oct

0924: Sep 24 08:11:20 - Sep 24 08:48:04  
0925: Sep 25 08:59:59 - Sep 25 09:31:29  
0926: Sep 26 02:37:36 - Sep 26 03:04:31  
0928: Sep 27 16:56:49 - Sep 28 07:36:58  
0929: Sep 29 02:59:52 - Sep 29 07:59:51  
0930: Sep 30 14:32:52 - Sep 30 15:54:50  
1001: Oct 01 05:10:57 - Oct 01 07:12:10  
1002: Oct 02 07:40:48 - Oct 02 08:47:00  
1003: Oct 03 04:43:20 - Oct 03 06:51:52  
1004: NA  
1005: Oct 05 03:19:38 - Oct 05 04:58:06  
1006: Oct 06 08:53:56 - Oct 06 09:55:46  
1007: Oct 07 05:44:31 - Oct 07 06:45:10  
1008: Oct 08 01:24:50 - Oct 08 02:42:36  
1009: Oct 09 01:56:22 - Oct 09 03:24:03  
1010: Oct 10 01:14:51 - Oct 10 06:54:25  
1011: Oct 11 12:24:45 - Oct 11 14:50:57  
1012: NA  
1013: NA  
1014: Oct 14 03:27:01 - Oct 14 14:12:38  
1015: Oct 15 05:06:51 - Oct 15 09:39:44  
1016: NA  
1017: NA  
1018: Oct 18 00:20:24 - Oct 18 05:44:20  
1019: Oct 19 04:47:38 - Oct 19 09:24:48  
1020: Oct 20 06:58:36 - Oct 20 12:02:31  
1021: Oct 21 05:53:39 - Oct 21 13:39:37  
1022: NA  
1023: Oct 23 02:42:46 - Oct 23 09:47:39  
1024: Oct 24 06:24:29 - Oct 24 13:14:49  
1025: Oct 25 00:18:05 - Oct 25 03:31:05

**how I obtained the merge data:**

0924: qhd Sep 24 08:11:20 - Sep 24 08:48:04  
0925: qhd Sep 25 08:59:59 - Sep 25 09:31:29  
0926: qhd Sep 26 02:37:36 - Sep 26 03:04:31  
0927: NA  
0928: qhd Sep 27 16:56:49 - Sep 28 07:36:58  
0929: qhd Sep 29 02:59:52 - Sep 29 07:59:51  
0930: jhy: Sep 30 09:58:44 - Sep 30 20:38:56  
1001: jhy Oct 01 08:17:38 - Oct 01 19:08:22  
1002: qhd Oct 02 07:40:48 - Oct 02 08:47:00  
1003: combine qhd + jhy  
1004: combine jhy + qhd  
1005: combine jhy + qhd  
1006 : combine jhy + qhd  
1007: combine jhy + qhd  
1008: combine qhd + jhy  
1009: combine qhd + jhy  
1010: qhd Oct 10 01:14:51 - Oct 10 06:54:25  
1011: qhd Oct 11 12:24:45 - Oct 11 14:50:57  
1012: extract from jhy 10  
1013: extract from jhy 10  
1014: qhd Oct 14 03:27:01 - Oct 14 14:12:38  
1015: qhd Oct 15 05:06:51 - Oct 15 09:39:44  
1016: jhy Oct 16 08:34:53 - Oct 16 12:36:27  
1017: jhy 13 extract  
1018: qhd Oct 18 00:20:24 - Oct 18 05:44:20  
1019: qhd Oct 19 04:47:38 - Oct 19 09:24:48  
1020: qhd Oct 20 06:58:36 - Oct 20 12:02:31  
1021: qhd Oct 21 05:53:39 - Oct 21 13:39:37  
1022: NA  
1023: qhd Oct 23 02:42:46 - Oct 23 09:47:39  
1024: qhd Oct 24 06:24:29 - Oct 24 13:14:49  
1025: qhd Oct 25 00:18:05 - Oct 25 03:31:05  
1026: jhy 16 extract  
1027: jhy 16 extract  
1028: jhy 16 extract  
1029: jhy 16 extract  
1030: jhy 16 extract