Interacciones en Twitter threads

Lectura de datos

Se leerán los 5 archivos de datos a través de pandas, se grafican los promedios y medianas.

In [93]:
```
import pandas as pd


csv1 = pd.read_csv('five_ten.csv', encoding='iso-8859-1')
csv2 = pd.read_csv('ten_fifteen.csv', encoding='iso-8859-1')
csv3 = pd.read_csv('fifteen_twenty.csv', encoding='iso-8859-1')
csv4 = pd.read_csv('twenty_twentyfive.csv', encoding='iso-8859-1')
csv5 = pd.read_csv('twentyfive_thirty.csv', encoding='iso-8859-1')
```

In [94]:
```
csv1.head()
```

Out[94]:

	id	thread_number	timestamp	text	retweets	likes	replies
0	999307110902050818	Thread 1	1527088356	Extraordinary evidence at Treasury committee f...	66	59	5
1	999307395712143360	Thread 1	1527088424	The Brexiter favourite Max Fac - would cost bu...	83	107	10
2	999307826265812992	Thread 1	1527088526	How does he arrive at the figure\r\r\n\r\r\n20...	6	11	2
3	999308153346052102	Thread 1	1527088604	Theresa May's New Customs Partnership is much ...	7	10	1
4	999308653894230022	Thread 1	1527088724	Mr Thompson said he did not expect the EU to r...	17	12	2

In [95]:
```
csv2.head()
```

Out[95]:

	id	thread_number	timestamp	text	retweets	likes	replies
0	972205190593089538	Thread 1	1520626754	Corbynâ??s speech to Scottish Labour conferenc...	711	837	62
1	972205328485048320	Thread 1	1520626787	Corbyn â??Make no mistake about it, reduced ac...	94	289	3
2	972205463105392645	Thread 1	1520626819	Corbyn: â??we will not accept an off the peg m...	65	217	3
3	972205538644758528	Thread 1	1520626837	Corbyn: â??the option of a new UK customs unio...	109	323	10
4	972205714113466370	Thread 1	1520626879	Corbyn wants a say on single mkt rules â??we c...	86	258	5

In [96]:
```
csv3.head()
```

Out[96]:

	id	thread_number	timestamp	text	retweets	likes	replies
0	998968203681427458	Thread 1	1527007554	1) WE HAVE A BREAKING FREAKING STORY HERE\r\r\...	335	679	94
1	998968303136727041	Thread 1	1527007578	2) I AM SO EXCITE I MIGHT NOT EVEN SWEAR OR GO...	60	294	13
2	998968508225589249	Thread 1	1527007626	3) CNN, AP, and MSNBC....	120	291	5
3	998968614509273088	Thread 1	1527007652	4) HAVE JUST BEEN RAIDED BY THE FCC https://t....	174	465	39
4	998969018781523975	Thread 1	1527007748	5) My one source inside one of these networks ...	242	558	18

In [97]:
```
csv4.head()
```

Out[97]:

	id	thread_number	timestamp	text	retweets	likes	replies
0	978656977705414657	Thread 1	1522164980	20 *more* reasons I voted to leave for #Brexit...	47	56	8
1	978656982755434497	Thread 1	1522164981	3. Member states have no right of initiative i...	9	18	2
2	978656980263948290	Thread 1	1522164981	2. It says it wants free and fair trade but th...	10	22	3
3	978656986228314112	Thread 1	1522164982	5. Virtue signalling as public policy. The EU ...	10	19	1
4	978656984370204673	Thread 1	1522164982	4. Its energy directives are directed by group...	9	19	2

In [98]:
```
csv5.head()
```

Out[98]:

	id	thread_number	timestamp	text	retweets	likes	replies
0	978874480968728576	Thread 1	1522216837	(1) @DeadofKnight68 asks a question. https://t...	402	620	33
1	978875352654188546	Thread 1	1522217045	(2) In North Korea, the Kim family and the Kor...	90	239	3
2	978875693722456064	Thread 1	1522217126	(3) The KPA needs a central figure who serves ...	59	216	2
3	978876717287772161	Thread 1	1522217370	(4) KPA generals live in palaces.\r\r\n\r\r\nS...	70	224	3
4	978876925161672705	Thread 1	1522217420	(5) The Kim family and the KPA are twin mafias...	85	238	2

Agrupar tweets por thread

Se agrupan los threads, cada thread es una columna y las filas son los tweets. Se agrupan los threads en 3 dataframes: retweets, likes y respuestas. Esto para cada dataset

In [99]:
```
csv1_grouped_by_thread = csv1.groupby(['thread_number'])
csv2_grouped_by_thread = csv2.groupby(['thread_number'])
csv3_grouped_by_thread = csv3.groupby(['thread_number'])
csv4_grouped_by_thread = csv4.groupby(['thread_number'])
csv5_grouped_by_thread = csv5.groupby(['thread_number'])
```

In [100]:
```
# retweets
retweets1 = {}
retweets2 = {}
retweets3 = {}
retweets4 = {}
retweets5 = {}

for thread, data in dict(list(csv1_grouped_by_thread)).items():
    retweets1[thread] = list(data['retweets'])

retweets_by_thread1 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in retweets1.items
()]))

for thread, data in dict(list(csv2_grouped_by_thread)).items():
    retweets2[thread] = list(data['retweets'])

retweets_by_thread2 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in retweets2.items
()]))

for thread, data in dict(list(csv3_grouped_by_thread)).items():
    retweets3[thread] = list(data['retweets'])

retweets_by_thread3 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in retweets3.items
()]))

for thread, data in dict(list(csv4_grouped_by_thread)).items():
    retweets4[thread] = list(data['retweets'])

retweets_by_thread4 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in retweets4.items
()]))

for thread, data in dict(list(csv5_grouped_by_thread)).items():
    retweets5[thread] = list(data['retweets'])

retweets_by_thread5 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in retweets5.items
()]))
```

In [101]:
```
retweets_by_thread1.head()
```

Out[101]:

	Thread 1	Thread 10	Thread 100	Thread 101	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	...	Thread 90	Thread 91
0	66.0	60.0	281.0	80.0	65	281.0	482.0	89.0	50.0	252.0	...	96.0	78.0
1	83.0	8.0	32.0	25.0	18	35.0	41.0	24.0	13.0	318.0	...	35.0	14.0
2	6.0	4.0	33.0	33.0	18	32.0	19.0	16.0	24.0	399.0	...	32.0	10.0
3	7.0	4.0	54.0	36.0	20	22.0	65.0	26.0	22.0	124.0	...	104.0	18.0
4	17.0	4.0	50.0	21.0	17	16.0	59.0	26.0	16.0	135.0	...	44.0	1.0

5 rows × 101 columns

In [102]:
```
retweets_by_thread2.head()
```

Out[102]:

	Thread 1	Thread 10	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	Thread 18	...	Thread 9	Thread 90
0	711.0	787.0	646.0	8.0	52.0	98.0	155.0	2109.0	17.0	13.0	...	1328	167.0
1	94.0	187.0	276.0	0.0	17.0	21.0	39.0	283.0	10.0	3.0	...	480	12.0
2	65.0	100.0	276.0	0.0	12.0	7.0	44.0	318.0	6.0	5.0	...	106	8.0
3	109.0	110.0	211.0	8.0	12.0	8.0	43.0	248.0	7.0	2.0	...	120	13.0
4	86.0	78.0	391.0	0.0	10.0	10.0	19.0	268.0	7.0	1.0	...	115	9.0

5 rows × 98 columns

In [103]:
```
retweets_by_thread3.head()
```

Out[103]:

	Thread 1	Thread 10	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	Thread 18	...	Thread 90	Thread 91
0	335.0	33	386	52.0	10.0	1.0	75.0	180.0	403	6.0	...	1625.0	25.0
1	60.0	6	65	17.0	0.0	0.0	1.0	8.0	124	1.0	...	89.0	845.0
2	120.0	5	96	13.0	0.0	0.0	2.0	6.0	179	0.0	...	82.0	13.0
3	174.0	3	100	15.0	0.0	0.0	3.0	7.0	95	1.0	...	89.0	3.0
4	242.0	3	73	11.0	0.0	0.0	3.0	6.0	104	1.0	...	90.0	4.0

5 rows × 99 columns

In [104]:
```
retweets_by_thread4.head()
```

Out[104]:

	Thread 1	Thread 10	Thread 100	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	...	Thread 90	Thread 91
0	47.0	201.0	919.0	25.0	6.0	45.0	90.0	14	1841.0	185.0	...	4	1211.0
1	9.0	43.0	136.0	16.0	0.0	4.0	9.0	3	234.0	25.0	...	10	160.0
2	10.0	25.0	131.0	22.0	0.0	4.0	8.0	1	258.0	45.0	...	2	119.0
3	10.0	22.0	103.0	61.0	1.0	6.0	10.0	1	226.0	36.0	...	2	140.0
4	9.0	23.0	168.0	43.0	3.0	6.0	8.0	1	189.0	40.0	...	1	96.0

5 rows × 100 columns

In [105]:
```
retweets_by_thread5.head()
```

Out[105]:

	Thread 1	Thread 10	Thread 100	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	...	Thread 90	Thread 91
0	402	36.0	18.0	5541.0	0.0	1523.0	412.0	13979.0	82.0	1525.0	...	301.0	30
1	90	10.0	9.0	1571.0	57.0	181.0	48.0	406.0	24.0	242.0	...	28.0	4
2	59	10.0	15.0	441.0	0.0	158.0	42.0	331.0	19.0	557.0	...	36.0	5
3	70	10.0	9.0	757.0	0.0	192.0	41.0	369.0	19.0	550.0	...	47.0	4
4	85	10.0	13.0	416.0	2.0	179.0	36.0	518.0	14.0	727.0	...	31.0	2

5 rows × 100 columns

In [106]:
```
# likes
likes1 = {}
likes2 = {}
likes3 = {}
likes4 = {}
likes5 = {}

for thread, data in dict(list(csv1_grouped_by_thread)).items():
    likes1[thread] = list(data['likes'])

likes_by_thread1 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in likes1.items()]))

for thread, data in dict(list(csv2_grouped_by_thread)).items():
    likes2[thread] = list(data['likes'])

likes_by_thread2 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in likes2.items()]))

for thread, data in dict(list(csv3_grouped_by_thread)).items():
    likes3[thread] = list(data['likes'])

likes_by_thread3 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in likes3.items()]))

for thread, data in dict(list(csv4_grouped_by_thread)).items():
    likes4[thread] = list(data['likes'])

likes_by_thread4 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in likes4.items()]))

for thread, data in dict(list(csv5_grouped_by_thread)).items():
    likes5[thread] = list(data['likes'])

likes_by_thread5 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in likes5.items()]))
```

In [107]:
```
likes_by_thread1.head()
```

Out[107]:

	Thread 1	Thread 10	Thread 100	Thread 101	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	...	Thread 90	Thread 91
0	59.0	156.0	382.0	115.0	78	327.0	524.0	116.0	68.0	438.0	...	230.0	64.0
1	107.0	39.0	59.0	39.0	35	49.0	154.0	35.0	27.0	402.0	...	61.0	14.0
2	11.0	30.0	70.0	40.0	35	46.0	92.0	27.0	46.0	587.0	...	80.0	13.0
3	10.0	32.0	94.0	60.0	49	37.0	190.0	36.0	38.0	208.0	...	177.0	26.0
4	12.0	26.0	62.0	46.0	30	20.0	156.0	42.0	22.0	223.0	...	95.0	3.0

5 rows × 101 columns

In [108]:
```
likes_by_thread2.head()
```

Out[108]:

	Thread 1	Thread 10	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	Thread 18	...	Thread 9	Thread 90
0	837.0	1787.0	992.0	24.0	44.0	273.0	254.0	2473.0	18.0	40.0	...	2406	172.0
1	289.0	437.0	464.0	2.0	14.0	62.0	90.0	513.0	10.0	8.0	...	1306	48.0
2	217.0	259.0	568.0	0.0	13.0	54.0	76.0	539.0	12.0	8.0	...	351	47.0
3	323.0	302.0	342.0	14.0	12.0	53.0	75.0	491.0	9.0	5.0	...	299	58.0
4	258.0	223.0	490.0	1.0	9.0	48.0	54.0	443.0	15.0	3.0	...	316	50.0

5 rows × 98 columns

In [109]:
```
likes_by_thread3.head()
```

Out[109]:

	Thread 1	Thread 10	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	Thread 18	...	Thread 90	Thread 91
0	679.0	37	459	73.0	15.0	9.0	97.0	147.0	730	13.0	...	2073.0	36.0
1	294.0	10	102	28.0	1.0	1.0	6.0	11.0	266	5.0	...	455.0	1091.0
2	291.0	21	128	17.0	1.0	1.0	5.0	10.0	323	3.0	...	383.0	16.0
3	465.0	10	127	21.0	1.0	1.0	8.0	18.0	215	3.0	...	448.0	7.0
4	558.0	14	100	16.0	2.0	2.0	7.0	11.0	213	3.0	...	468.0	10.0

5 rows × 99 columns

In [110]:
```
likes_by_thread4.head()
```

Out[110]:

	Thread 1	Thread 10	Thread 100	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	...	Thread 90	Thread 91
0	56.0	151.0	2184.0	25.0	49.0	53.0	166.0	18	4017.0	203.0	...	3	1892.0
1	18.0	26.0	505.0	13.0	1.0	11.0	29.0	3	654.0	46.0	...	10	604.0
2	22.0	22.0	457.0	16.0	1.0	12.0	28.0	2	551.0	67.0	...	2	507.0
3	19.0	16.0	316.0	54.0	5.0	12.0	29.0	2	692.0	61.0	...	1	554.0
4	19.0	15.0	443.0	71.0	8.0	13.0	26.0	2	530.0	61.0	...	1	459.0

5 rows × 100 columns

In [111]:
```
likes_by_thread5.head()
```

Out[111]:

	Thread 1	Thread 10	Thread 100	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	...	Thread 90	Thread 91
0	620	34.0	35.0	6489.0	8.0	3302.0	576.0	17285.0	115.0	2461.0	...	475.0	58
1	239	3.0	34.0	2004.0	101.0	649.0	186.0	2780.0	44.0	465.0	...	94.0	23
2	216	3.0	49.0	883.0	12.0	529.0	164.0	2753.0	43.0	920.0	...	127.0	29
3	224	3.0	43.0	1307.0	17.0	524.0	163.0	2648.0	38.0	1153.0	...	119.0	27
4	238	2.0	38.0	841.0	43.0	690.0	173.0	3734.0	32.0	1089.0	...	112.0	22

5 rows × 100 columns

In [112]:
```
# replies

replies1 = {}
replies2 = {}
replies3 = {}
replies4 = {}
replies5 = {}

for thread, data in dict(list(csv1_grouped_by_thread)).items():
    replies1[thread] = list(data['replies'])

replies_by_thread1 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in replies1.items
()]))

for thread, data in dict(list(csv2_grouped_by_thread)).items():
    replies2[thread] = list(data['replies'])

replies_by_thread2 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in replies2.items
()]))

for thread, data in dict(list(csv3_grouped_by_thread)).items():
    replies3[thread] = list(data['replies'])

replies_by_thread3 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in replies3.items
()]))

for thread, data in dict(list(csv4_grouped_by_thread)).items():
    replies4[thread] = list(data['replies'])

replies_by_thread4 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in replies4.items
()]))

for thread, data in dict(list(csv5_grouped_by_thread)).items():
    replies5[thread] = list(data['replies'])

replies_by_thread5 = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in replies5.items
()]))
```

In [113]:
```
replies_by_thread1.head()
```

Out[113]:

	Thread 1	Thread 10	Thread 100	Thread 101	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	...	Thread 90	Thread 91
0	5.0	18.0	17.0	9.0	7	14.0	41.0	14.0	4.0	74.0	...	13.0	4.0
1	10.0	2.0	4.0	1.0	3	1.0	6.0	1.0	3.0	34.0	...	1.0	1.0
2	2.0	1.0	2.0	2.0	2	1.0	1.0	1.0	5.0	35.0	...	2.0	2.0
3	1.0	2.0	8.0	3.0	2	3.0	1.0	3.0	4.0	39.0	...	1.0	3.0
4	2.0	2.0	1.0	5.0	1	1.0	4.0	4.0	4.0	17.0	...	3.0	1.0

5 rows × 101 columns

In [114]:
```
replies_by_thread2.head()
```

Out[114]:

	Thread 1	Thread 10	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	Thread 18	...	Thread 9	Thread 90
0	62.0	78.0	66.0	1.0	5.0	18.0	12.0	100.0	1.0	4.0	...	75	9.0
1	3.0	19.0	15.0	1.0	1.0	5.0	10.0	9.0	2.0	1.0	...	0	1.0
2	3.0	7.0	79.0	1.0	1.0	5.0	7.0	9.0	2.0	1.0	...	3	1.0
3	10.0	13.0	16.0	2.0	1.0	2.0	1.0	5.0	2.0	1.0	...	7	1.0
4	5.0	1.0	58.0	1.0	1.0	3.0	1.0	13.0	1.0	1.0	...	5	2.0

5 rows × 98 columns

In [115]:
```
replies_by_thread3.head()
```

Out[115]:

	Thread 1	Thread 10	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	Thread 18	...	Thread 90	Thread 91
0	94.0	7	23	2.0	1.0	2.0	13.0	12.0	31	2.0	...	127.0	16.0
1	13.0	2	1	4.0	1.0	1.0	1.0	1.0	8	1.0	...	9.0	1.0
2	5.0	1	7	1.0	1.0	1.0	1.0	2.0	15	1.0	...	4.0	1.0
3	39.0	1	4	2.0	1.0	1.0	1.0	2.0	5	1.0	...	4.0	1.0
4	18.0	1	4	1.0	1.0	1.0	1.0	2.0	4	1.0	...	6.0	1.0

5 rows × 99 columns

In [116]:
```
replies_by_thread4.head()
```

Out[116]:

	Thread 1	Thread 10	Thread 100	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	...	Thread 90	Thread 91
0	8.0	5.0	127.0	7.0	4.0	5.0	5.0	1	138.0	22.0	...	1	85.0
1	2.0	2.0	15.0	3.0	1.0	2.0	1.0	1	14.0	1.0	...	1	21.0
2	3.0	2.0	71.0	2.0	1.0	1.0	1.0	1	9.0	6.0	...	1	45.0
3	1.0	2.0	4.0	54.0	1.0	1.0	4.0	1	19.0	2.0	...	1	14.0
4	2.0	3.0	20.0	7.0	2.0	1.0	1.0	1	7.0	1.0	...	1	7.0

5 rows × 100 columns

In [117]:
```
replies_by_thread5.head()
```

Out[117]:

	Thread 1	Thread 10	Thread 100	Thread 11	Thread 12	Thread 13	Thread 14	Thread 15	Thread 16	Thread 17	...	Thread 90	Thread 91
0	33	4.0	2.0	341.0	6.0	193.0	69.0	382.0	7.0	98.0	...	55.0	8
1	3	1.0	1.0	0.0	1.0	2.0	8.0	0.0	1.0	8.0	...	3.0	1
2	2	3.0	3.0	0.0	1.0	12.0	5.0	0.0	1.0	23.0	...	2.0	2
3	3	1.0	1.0	7.0	1.0	2.0	2.0	0.0	1.0	0.0	...	4.0	2
4	2	2.0	1.0	2.0	1.0	16.0	4.0	0.0	1.0	0.0	...	4.0	1

5 rows × 100 columns

Largo promedio por threads.

In [118]:
```
average_length1 = csv1_grouped_by_thread.size().mean()
print(average_length1)
average_length2 = csv2_grouped_by_thread.size().mean()
print(average_length2)
average_length3 = csv3_grouped_by_thread.size().mean()
print(average_length3)
average_length4 = csv4_grouped_by_thread.size().mean()
print(average_length4)
average_length5 = csv5_grouped_by_thread.size().mean()
print(average_length5)
```

```
7.98019801980198
12.73469387755102
17.666666666666668
22.96
27.9
```

Gráficos de las reacciones promedios por número de tweet.

In [119]:
```
from bokeh.io import output_notebook, show
from bokeh.plotting import figure
from bokeh.models import Span

output_notebook()
```

(https://bokeh.pydata.org)

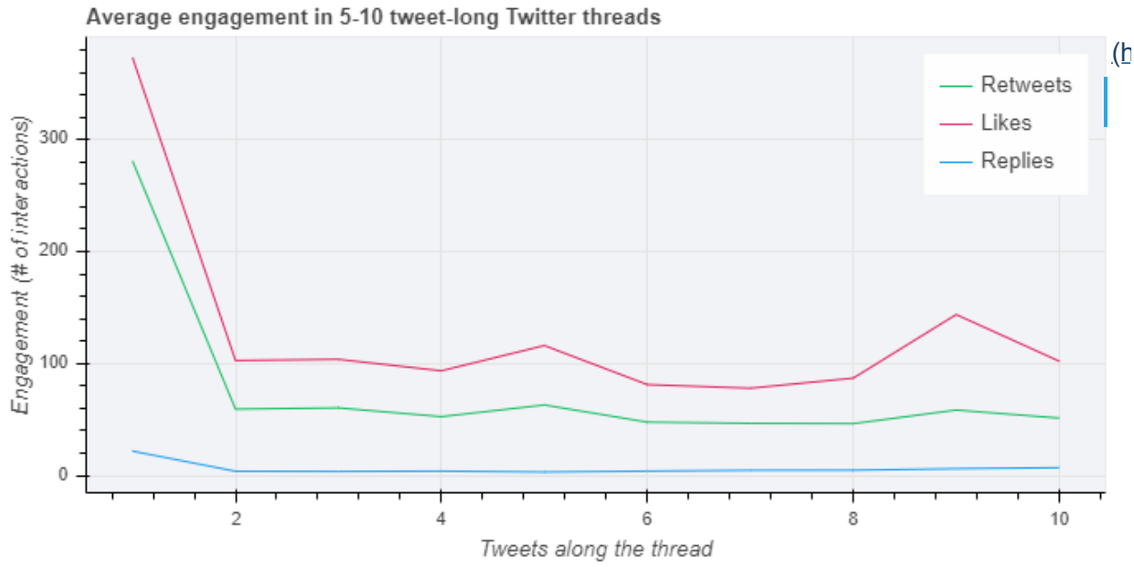
Bokeh JS 1.0.4 successfully loaded.

In [120]:
```
# averages
avg1 = pd.DataFrame()
avg1['Retweets'] = retweets_by_thread1.mean(axis=1)
avg1['Likes'] = likes_by_thread1.mean(axis=1)
avg1['Replies'] = replies_by_thread1.mean(axis=1)

average_engagement1 = figure(plot_width=700,
           plot_height=350,
           title='Average engagement in 5-10 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

average_engagement1.line(list(range(1,11)), avg1['Retweets'].values,line_color='#17bf63'
, legend='Retweets')
average_engagement1.line(list(range(1,11)), avg1['Likes'].values, line_color='#e0245e',
legend='Likes')
average_engagement1.line(list(range(1,11)), avg1['Replies'].values, line_color='#1da1f2'
, legend='Replies')

show(average_engagement1)
```



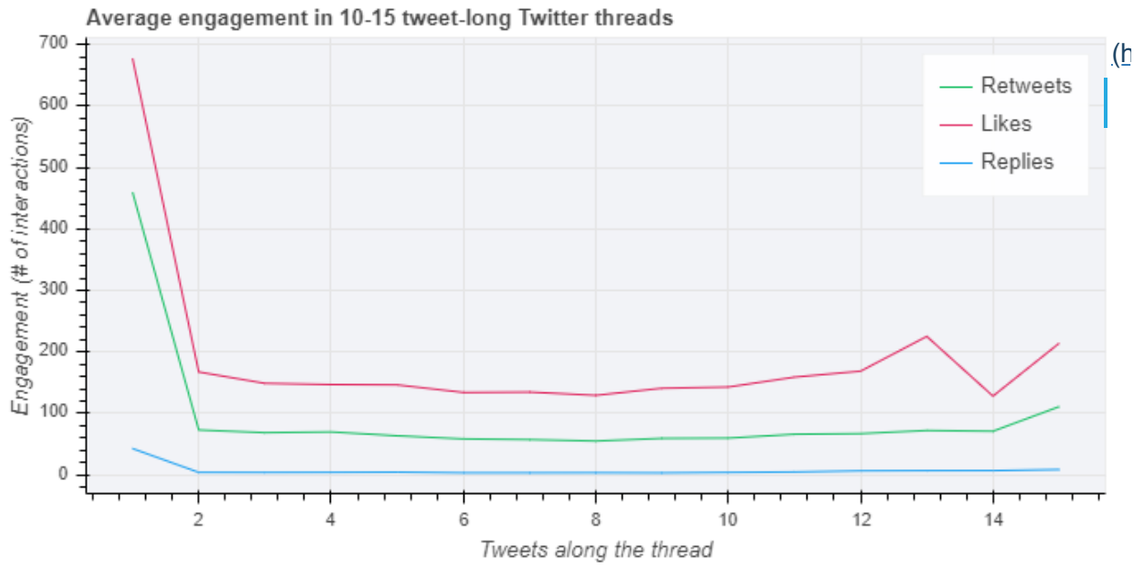
In [121]:

```
# averages
avg2 = pd.DataFrame()
avg2['Retweets'] = retweets_by_thread2.mean(axis=1)
avg2['Likes'] = likes_by_thread2.mean(axis=1)
avg2['Replies'] = replies_by_thread2.mean(axis=1)

average_engagement2 = figure(plot_width=700,
            plot_height=350,
            title='Average engagement in 10-15 tweet-long Twitter threads',
            background_fill_color="#f2f3f7",
            y_axis_label='Engagement (# of interactions)',
            x_axis_label='Tweets along the thread')

average_engagement2.line(list(range(1,16)), avg2['Retweets'].values,line_color='#17bf63'
, legend='Retweets')
average_engagement2.line(list(range(1,16)), avg2['Likes'].values, line_color='#e0245e',
legend='Likes')
average_engagement2.line(list(range(1,16)), avg2['Replies'].values, line_color='#1da1f2'
, legend='Replies')

show(average_engagement2)
```



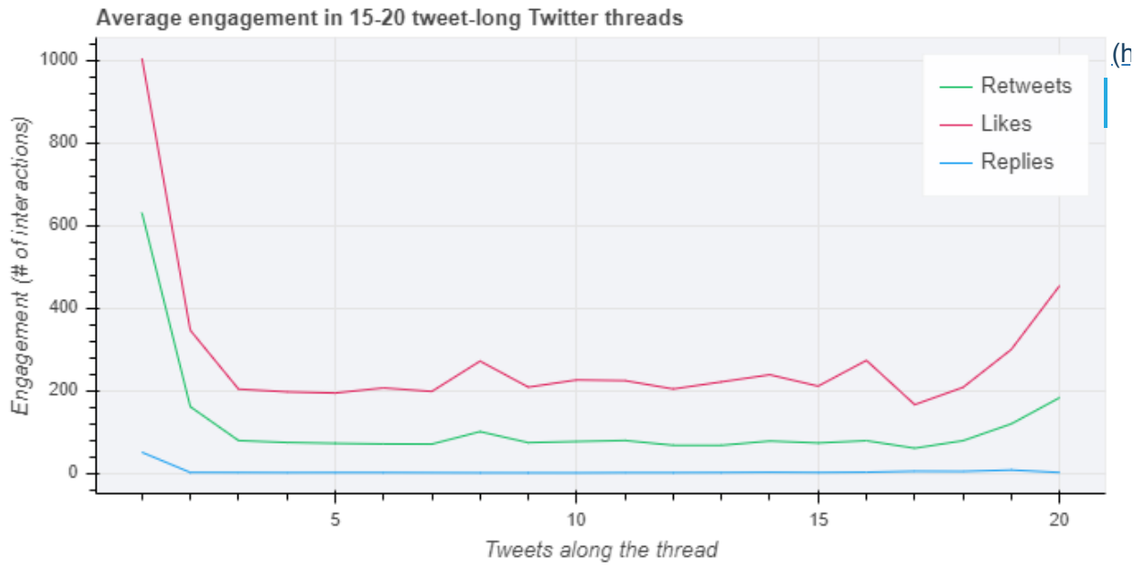
In [122]:

```
# averages
avg3 = pd.DataFrame()
avg3['Retweets'] = retweets_by_thread3.mean(axis=1)
avg3['Likes'] = likes_by_thread3.mean(axis=1)
avg3['Replies'] = replies_by_thread3.mean(axis=1)

average_engagement3 = figure(plot_width=700,
           plot_height=350,
           title='Average engagement in 15-20 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

average_engagement3.line(list(range(1,21)), avg3['Retweets'].values,line_color='#17bf63'
, legend='Retweets')
average_engagement3.line(list(range(1,21)), avg3['Likes'].values, line_color='#e0245e',
legend='Likes')
average_engagement3.line(list(range(1,21)), avg3['Replies'].values, line_color='#1da1f2'
, legend='Replies')

show(average_engagement3)
```



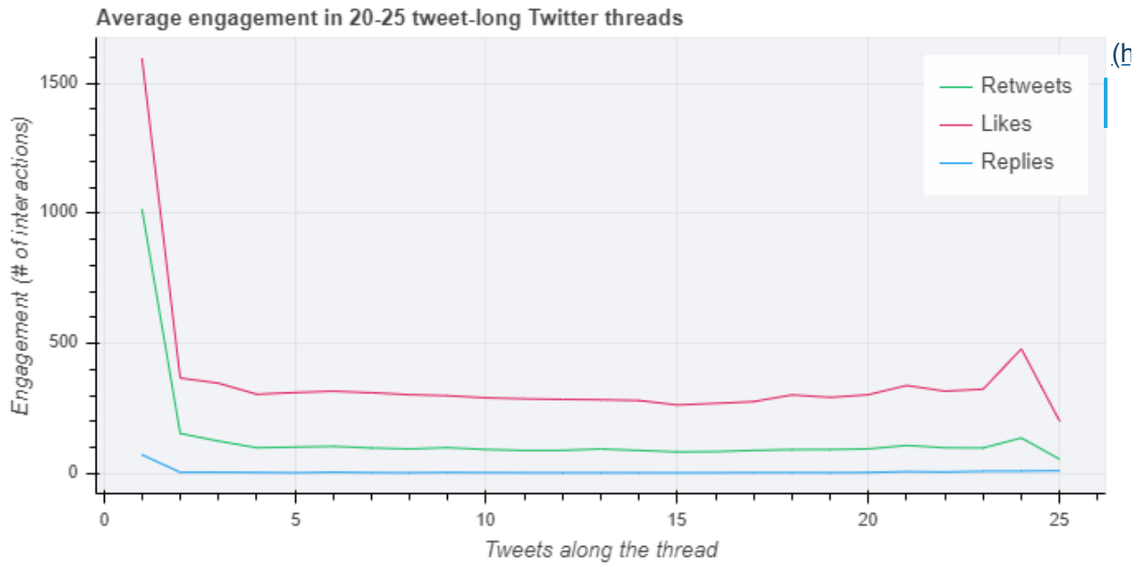
In [123]:

```
# averages
avg4 = pd.DataFrame()
avg4['Retweets'] = retweets_by_thread4.mean(axis=1)
avg4['Likes'] = likes_by_thread4.mean(axis=1)
avg4['Replies'] = replies_by_thread4.mean(axis=1)

average_engagement4 = figure(plot_width=700,
           plot_height=350,
           title='Average engagement in 20-25 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

average_engagement4.line(list(range(1,26)), avg4['Retweets'].values,line_color='#17bf63'
, legend='Retweets')
average_engagement4.line(list(range(1,26)), avg4['Likes'].values, line_color='#e0245e',
legend='Likes')
average_engagement4.line(list(range(1,26)), avg4['Replies'].values, line_color='#1da1f2'
, legend='Replies')

show(average_engagement4)
```

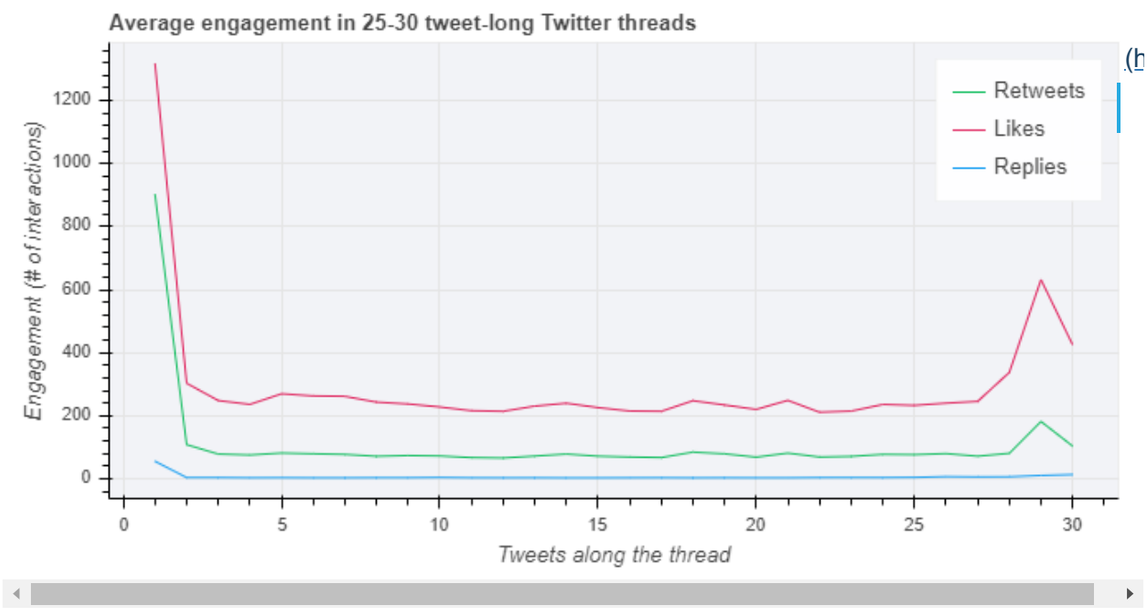


In [124]:
```
# averages
avg5 = pd.DataFrame()
avg5['Retweets'] = retweets_by_thread5.mean(axis=1)
avg5['Likes'] = likes_by_thread5.mean(axis=1)
avg5['Replies'] = replies_by_thread5.mean(axis=1)

average_engagement5 = figure(plot_width=700,
           plot_height=350,
           title='Average engagement in 25-30 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

average_engagement5.line(list(range(1,31)), avg5['Retweets'].values,line_color='#17bf63'
, legend='Retweets')
average_engagement5.line(list(range(1,31)), avg5['Likes'].values, line_color='#e0245e',
legend='Likes')
average_engagement5.line(list(range(1,31)), avg5['Replies'].values, line_color='#1da1f2'
, legend='Replies')

show(average_engagement5)
```



Análisis de distribución de reacciones por tweet

In [125]:
```
from bokeh.layouts import row as bokeh_row
```

In [126]:

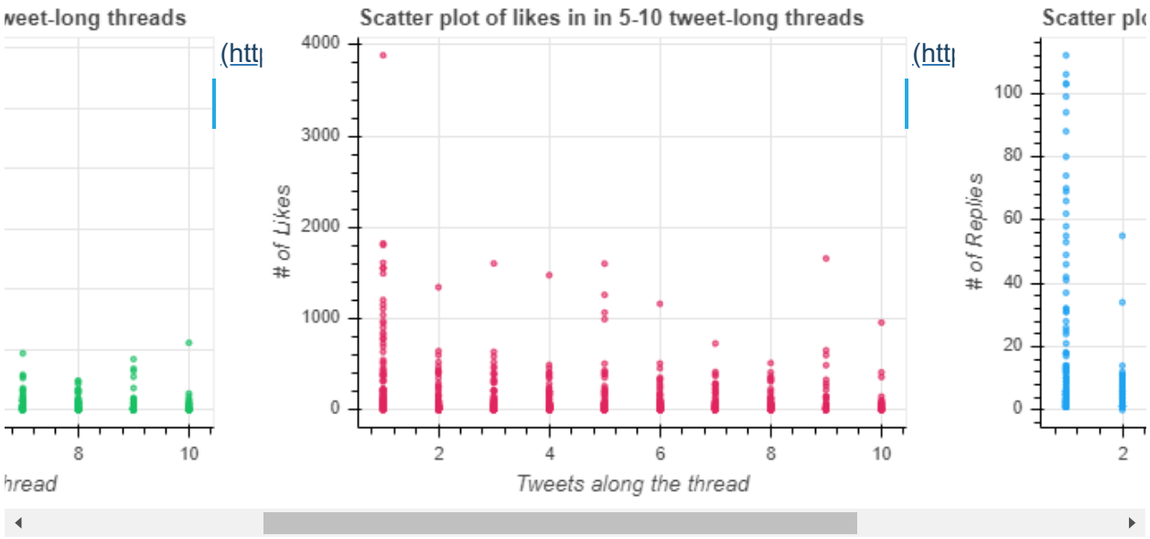
```
scatter_rts1 = figure(plot_width=420, plot_height=310, title='Scatter plot of retweets i
n 5-10 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='# of R
etweets')
scatter_likes1 = figure(plot_width=420, plot_height=310, title='Scatter plot of likes in
in 5-10 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='# of
 Likes')
scatter_replies1 = figure(plot_width=420, plot_height=310, title='Scatter plot of replie
s in 5-10 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='# o
f Replies')

# add each data point to the retweets scatter plot
for row in retweets_by_thread1:
    scatter_rts1.circle(list(range(1,11)), retweets_by_thread1.loc[:, row], size=3, line
_color="#17bf63", fill_color="#17bf63", fill_alpha=0.5)

# add each data point to the likes scatter plot
for row in likes_by_thread1:
    scatter_likes1.circle(list(range(1,11)), likes_by_thread1.loc[:, row], size=3, line_
color="#e0245e", fill_color="#e0245e", fill_alpha=0.5)

# add each data point to the replies scatter plot
for row in replies_by_thread1:
    scatter_replies1.circle(list(range(1,11)), replies_by_thread1.loc[:, row], size=3, l
ine_color="#1da1f2", fill_color="#1da1f2", fill_alpha=0.5)

show(bokeh_row(scatter_rts1, scatter_likes1, scatter_replies1))
```



In [127]:

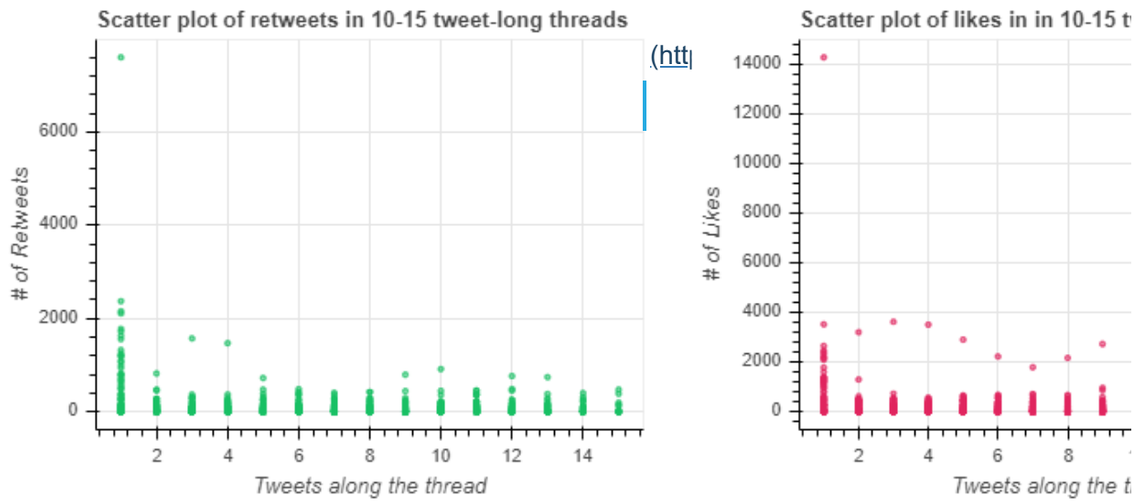
```
scatter_rts2 = figure(plot_width=420, plot_height=310, title='Scatter plot of retweets i
n 10-15 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='# of
 Retweets')
scatter_likes2 = figure(plot_width=420, plot_height=310, title='Scatter plot of likes in
in 10-15 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='# of
Likes')
scatter_replies2 = figure(plot_width=420, plot_height=310, title='Scatter plot of replie
s in 10-15 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='#
 of Replies')

# add each data point to the retweets scatter plot
for row in retweets_by_thread2:
    scatter_rts2.circle(list(range(1,16)), retweets_by_thread2.loc[:, row], size=3, line
_color="#17bf63", fill_color="#17bf63", fill_alpha=0.5)

# add each data point to the likes scatter plot
for row in likes_by_thread2:
    scatter_likes2.circle(list(range(1,16)), likes_by_thread2.loc[:, row], size=3, line_
color="#e0245e", fill_color="#e0245e", fill_alpha=0.5)

# add each data point to the replies scatter plot
for row in replies_by_thread2:
    scatter_replies2.circle(list(range(1,16)), replies_by_thread2.loc[:, row], size=3, l
ine_color="#1da1f2", fill_color="#1da1f2", fill_alpha=0.5)

show(bokeh_row(scatter_rts2, scatter_likes2, scatter_replies2))
```



In [128]:

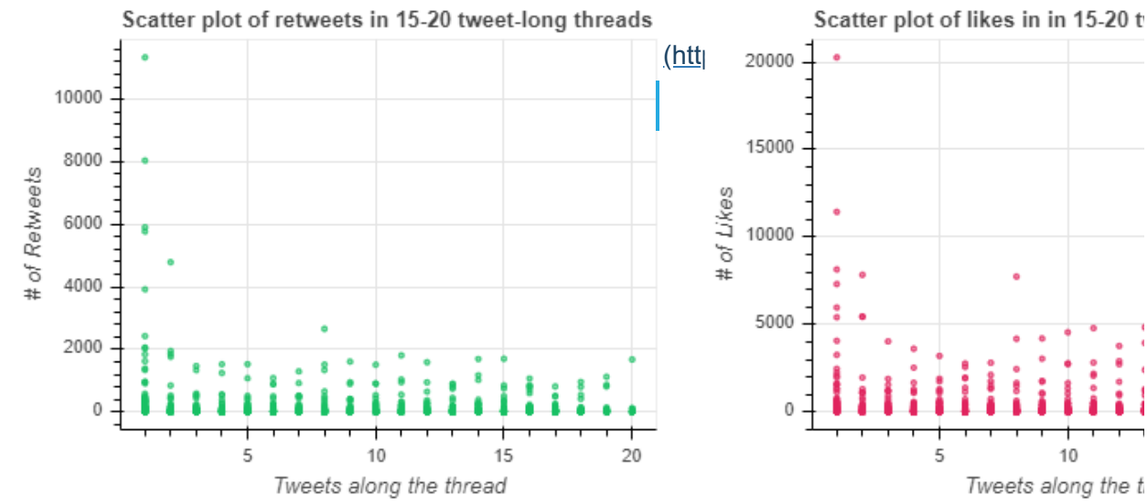
```
scatter_rts3 = figure(plot_width=420, plot_height=310, title='Scatter plot of retweets i
n 15-20 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='# of
 Retweets')
scatter_likes3 = figure(plot_width=420, plot_height=310, title='Scatter plot of likes in
in 15-20 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='# of
Likes')
scatter_replies3 = figure(plot_width=420, plot_height=310, title='Scatter plot of replie
s in 15-20 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='#
 of Replies')

# add each data point to the retweets scatter plot
for row in retweets_by_thread3:
    scatter_rts3.circle(list(range(1,21)), retweets_by_thread3.loc[:, row], size=3, line
_color="#17bf63", fill_color="#17bf63", fill_alpha=0.5)

# add each data point to the likes scatter plot
for row in likes_by_thread3:
    scatter_likes3.circle(list(range(1,21)), likes_by_thread3.loc[:, row], size=3, line_
color="#e0245e", fill_color="#e0245e", fill_alpha=0.5)

# add each data point to the replies scatter plot
for row in replies_by_thread3:
    scatter_replies3.circle(list(range(1,21)), replies_by_thread3.loc[:, row], size=3, l
ine_color="#1da1f2", fill_color="#1da1f2", fill_alpha=0.5)

show(bokeh_row(scatter_rts3, scatter_likes3, scatter_replies3))
```



In [129]:

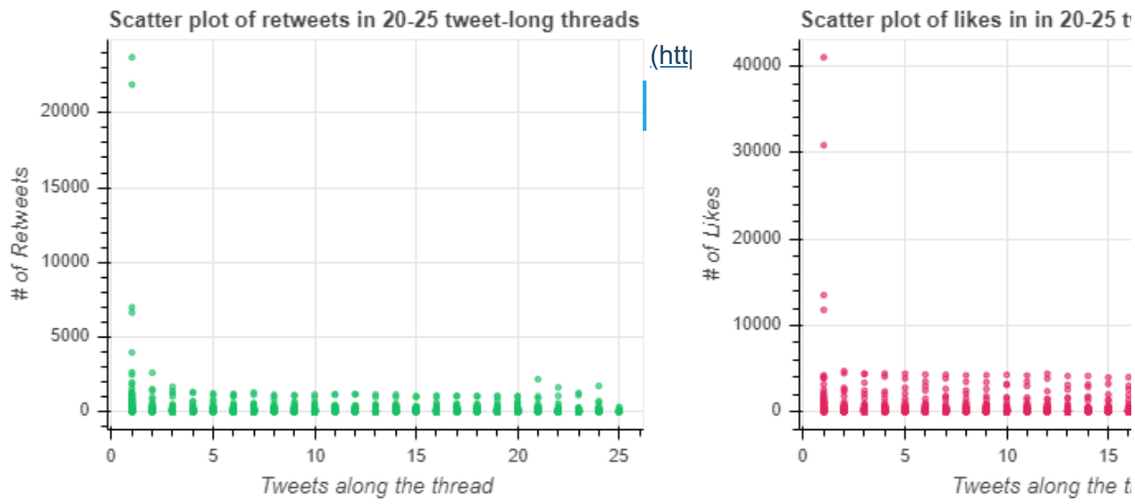
```
scatter_rts4 = figure(plot_width=420, plot_height=310, title='Scatter plot of retweets i
n 20-25 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='# of
 Retweets')
scatter_likes4 = figure(plot_width=420, plot_height=310, title='Scatter plot of likes in
in 20-25 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='# of
Likes')
scatter_replies4 = figure(plot_width=420, plot_height=310, title='Scatter plot of replie
s in 20-25 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='#
 of Replies')

# add each data point to the retweets scatter plot
for row in retweets_by_thread4:
    scatter_rts4.circle(list(range(1,26)), retweets_by_thread4.loc[:, row], size=3, line
_color="#17bf63", fill_color="#17bf63", fill_alpha=0.5)

# add each data point to the likes scatter plot
for row in likes_by_thread4:
    scatter_likes4.circle(list(range(1,26)), likes_by_thread4.loc[:, row], size=3, line_
color="#e0245e", fill_color="#e0245e", fill_alpha=0.5)

# add each data point to the replies scatter plot
for row in replies_by_thread4:
    scatter_replies4.circle(list(range(1,26)), replies_by_thread4.loc[:, row], size=3, l
ine_color="#1da1f2", fill_color="#1da1f2", fill_alpha=0.5)

show(bokeh_row(scatter_rts4, scatter_likes4, scatter_replies4))
```



In [130]:
```
scatter_rts5 = figure(plot_width=420, plot_height=310, title='Scatter plot of retweets i
n 25-30 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='# of
 Retweets')
scatter_likes5 = figure(plot_width=420, plot_height=310, title='Scatter plot of likes in
in 25-30 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='# of
Likes')
scatter_replies5 = figure(plot_width=420, plot_height=310, title='Scatter plot of replie
s in 25-30 tweet-long threads', x_axis_label='Tweets along the thread', y_axis_label='#
 of Replies')

# add each data point to the retweets scatter plot
for row in retweets_by_thread5:
    scatter_rts5.circle(list(range(1,31)), retweets_by_thread5.loc[:, row], size=3, line
_color="#17bf63", fill_color="#17bf63", fill_alpha=0.5)

# add each data point to the likes scatter plot
for row in likes_by_thread5:
    scatter_likes5.circle(list(range(1,31)), likes_by_thread5.loc[:, row], size=3, line_
color="#e0245e", fill_color="#e0245e", fill_alpha=0.5)

# add each data point to the replies scatter plot
for row in replies_by_thread5:
    scatter_replies5.circle(list(range(1,31)), replies_by_thread5.loc[:, row], size=3, l
ine_color="#1da1f2", fill_color="#1da1f2", fill_alpha=0.5)

show(bokeh_row(scatter_rts5, scatter_likes5, scatter_replies5))
```

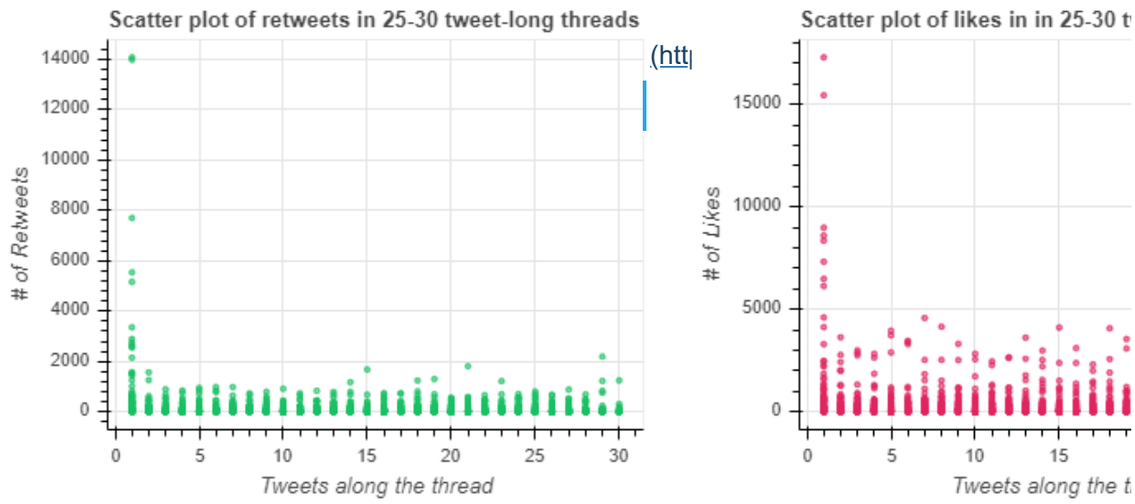


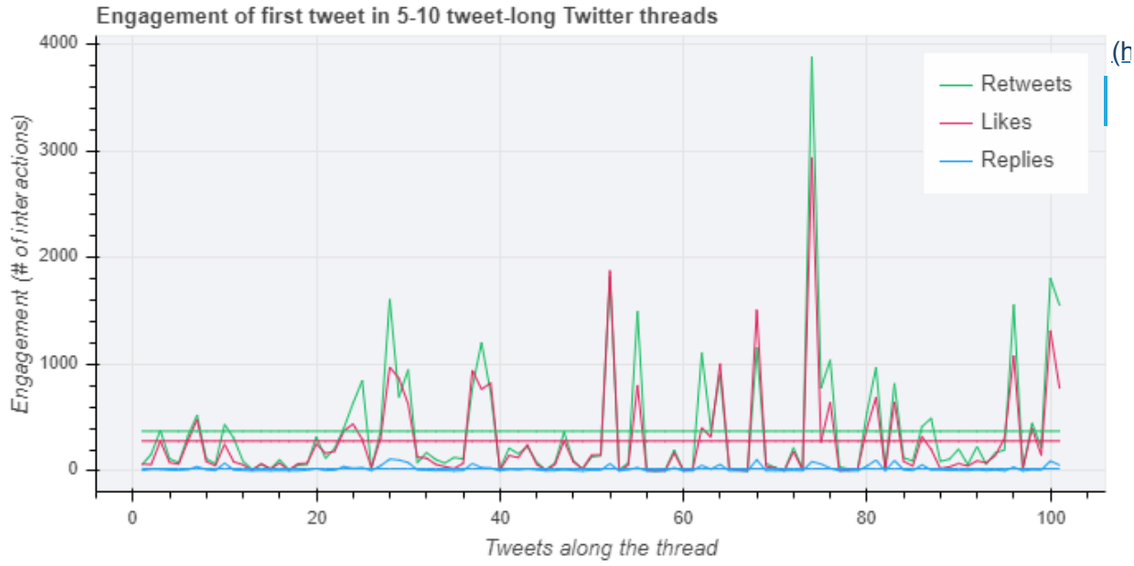
Gráfico reacciones primer tweet

In [131]:
```
# first tweet
engagement1 = figure(plot_width=700,
           plot_height=350,
           title='Engagement of first tweet in 5-10 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

engagement1.line(list(range(1,102)), likes_by_thread1.iloc[0],line_color='#17bf63', lege
nd='Retweets')
engagement1.line(list(range(1,102)), retweets_by_thread1.iloc[0].values, line_color='#e0
245e', legend='Likes')
engagement1.line(list(range(1,102)), replies_by_thread1.iloc[0].values, line_color='#1da
1f2', legend='Replies')
engagement1.line(list(range(1,102)), likes_by_thread1.iloc[0].mean(),line_color='#17bf6
3', legend='Retweets')
engagement1.line(list(range(1,102)), retweets_by_thread1.iloc[0].mean(), line_color='#e0
245e', legend='Likes')
engagement1.line(list(range(1,102)), replies_by_thread1.iloc[0].mean(), line_color='#1da
1f2', legend='Replies')



show(engagement1)
```



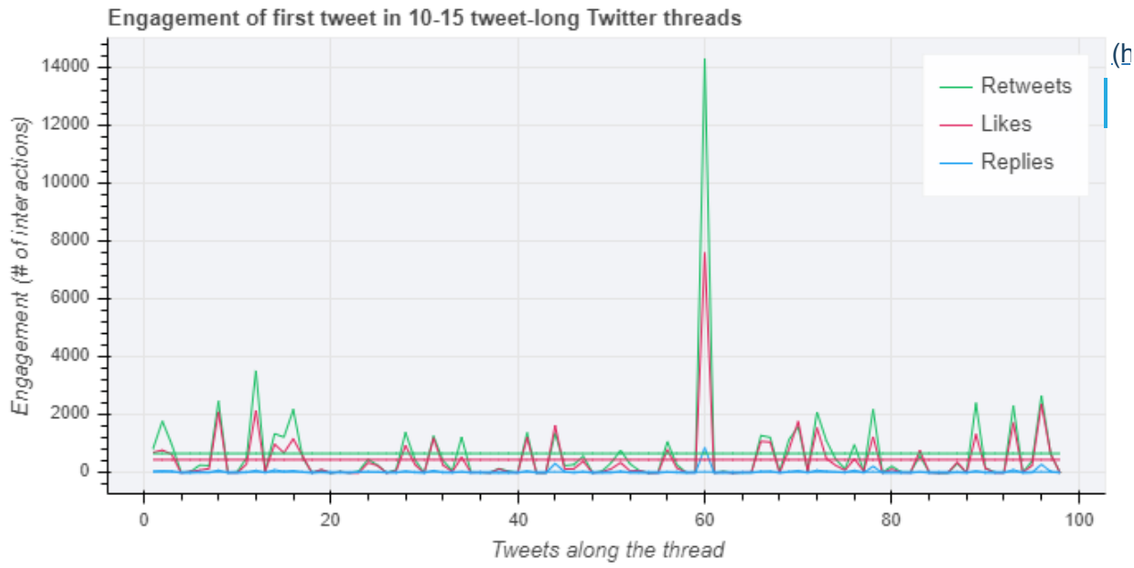
In [132]:

```
# first tweet
engagement2 = figure(plot_width=700,
           plot_height=350,
           title='Engagement of first tweet in 10-15 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

engagement2.line(list(range(1,99)), likes_by_thread2.iloc[0],line_color='#17bf63', legend='Retweets')
engagement2.line(list(range(1,99)), retweets_by_thread2.iloc[0].values, line_color='#e0245e', legend='Likes')
engagement2.line(list(range(1,99)), replies_by_thread2.iloc[0].values, line_color='#1da1f2', legend='Replies')
engagement2.line(list(range(1,99)), likes_by_thread2.iloc[0].mean(),line_color='#17bf63', legend='Retweets')
engagement2.line(list(range(1,99)), retweets_by_thread2.iloc[0].mean(), line_color='#e0245e', legend='Likes')
engagement2.line(list(range(1,99)), replies_by_thread2.iloc[0].mean(), line_color='#1da1f2', legend='Replies')



show(engagement2)
```

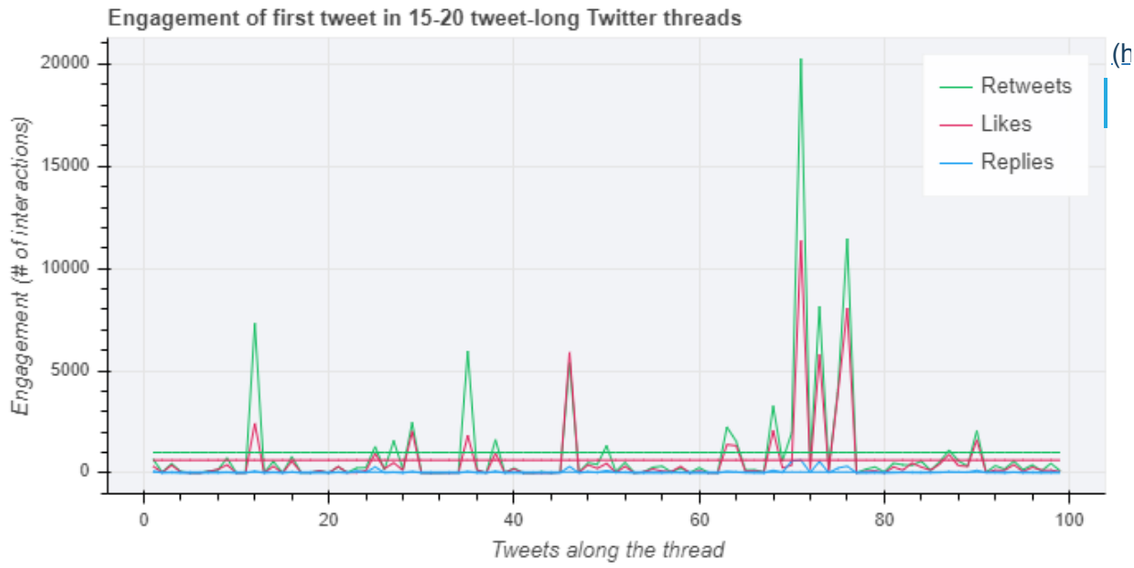


In [133]:
```
# first tweet
engagement3 = figure(plot_width=700,
           plot_height=350,
           title='Engagement of first tweet in 15-20 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

engagement3.line(list(range(1,100)), likes_by_thread3.iloc[0],line_color='#17bf63', lege
nd='Retweets')
engagement3.line(list(range(1,100)), retweets_by_thread3.iloc[0].values, line_color='#e0
245e', legend='Likes')
engagement3.line(list(range(1,100)), replies_by_thread3.iloc[0].values, line_color='#1da
1f2', legend='Replies')
engagement3.line(list(range(1,100)), likes_by_thread3.iloc[0].mean(),line_color='#17bf6
3', legend='Retweets')
engagement3.line(list(range(1,100)), retweets_by_thread3.iloc[0].mean(), line_color='#e0
245e', legend='Likes')
engagement3.line(list(range(1,100)), replies_by_thread3.iloc[0].mean(), line_color='#1da
1f2', legend='Replies')



show(engagement3)
```



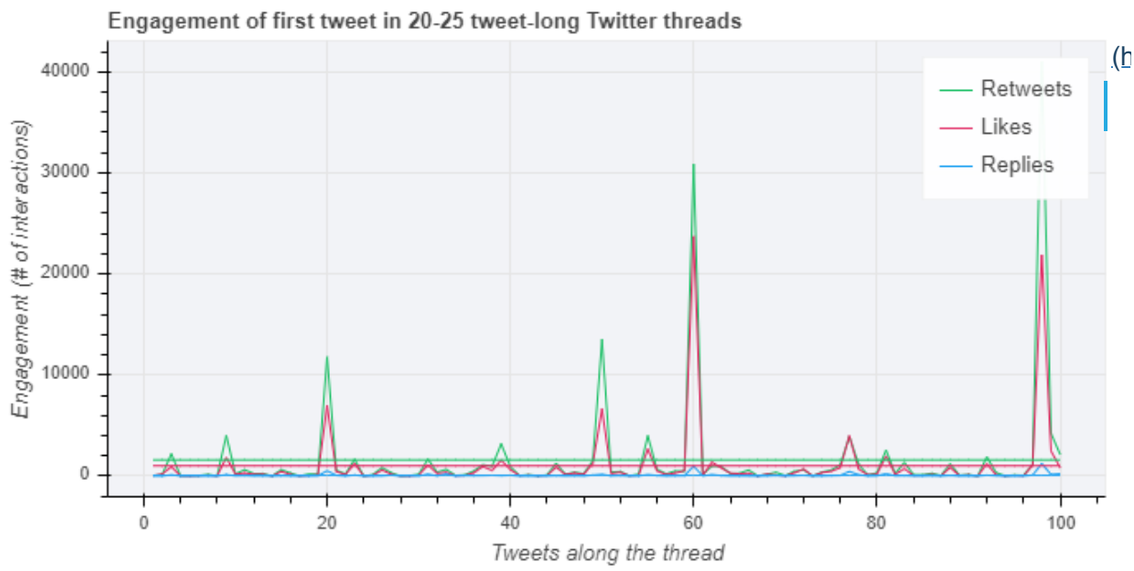
In [134]:

```
# first tweet
engagement4 = figure(plot_width=700,
           plot_height=350,
           title='Engagement of first tweet in 20-25 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

engagement4.line(list(range(1,101)), likes_by_thread4.iloc[0],line_color='#17bf63', lege
nd='Retweets')
engagement4.line(list(range(1,101)), retweets_by_thread4.iloc[0].values, line_color='#e0
245e', legend='Likes')
engagement4.line(list(range(1,101)), replies_by_thread4.iloc[0].values, line_color='#1da
1f2', legend='Replies')
engagement4.line(list(range(1,101)), likes_by_thread4.iloc[0].mean(),line_color='#17bf6
3', legend='Retweets')
engagement4.line(list(range(1,101)), retweets_by_thread4.iloc[0].mean(), line_color='#e0
245e', legend='Likes')
engagement4.line(list(range(1,101)), replies_by_thread4.iloc[0].mean(), line_color='#1da
1f2', legend='Replies')



show(engagement4)
```

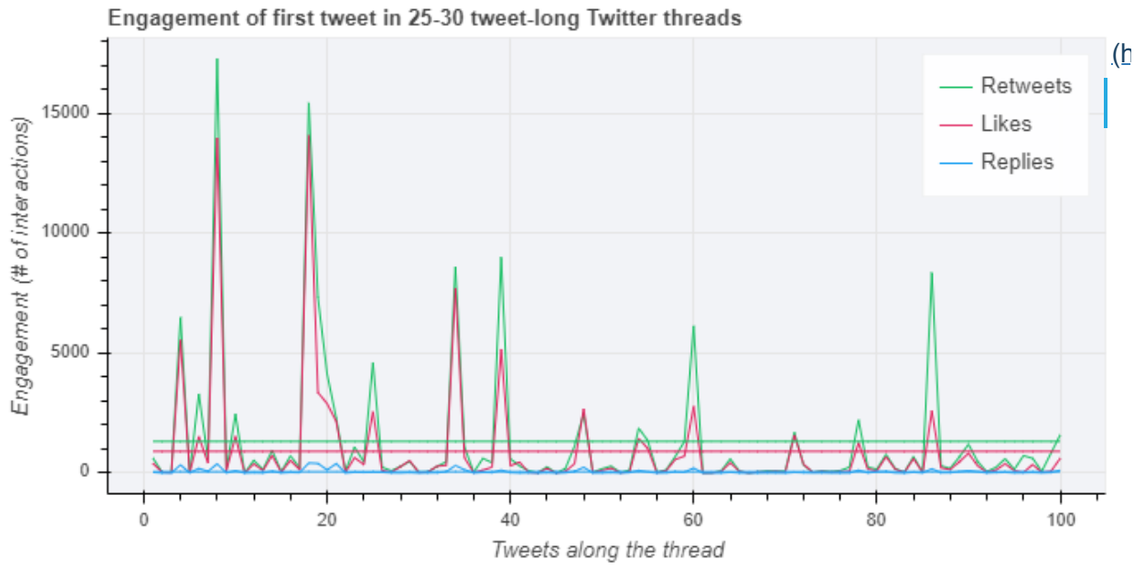


In [135]:
```
# first tweet
engagement5 = figure(plot_width=700,
           plot_height=350,
           title='Engagement of first tweet in 25-30 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

engagement5.line(list(range(1,101)), likes_by_thread5.iloc[0],line_color='#17bf63', lege
nd='Retweets')
engagement5.line(list(range(1,101)), retweets_by_thread5.iloc[0].values, line_color='#e0
245e', legend='Likes')
engagement5.line(list(range(1,101)), replies_by_thread5.iloc[0].values, line_color='#1da
1f2', legend='Replies')
engagement5.line(list(range(1,101)), likes_by_thread5.iloc[0].mean(),line_color='#17bf6
3', legend='Retweets')
engagement5.line(list(range(1,101)), retweets_by_thread5.iloc[0].mean(), line_color='#e0
245e', legend='Likes')
engagement5.line(list(range(1,101)), replies_by_thread5.iloc[0].mean(), line_color='#1da
1f2', legend='Replies')



show(engagement5)
```



Mediana por tweet

Let's make more graphs to better understand what's happening...

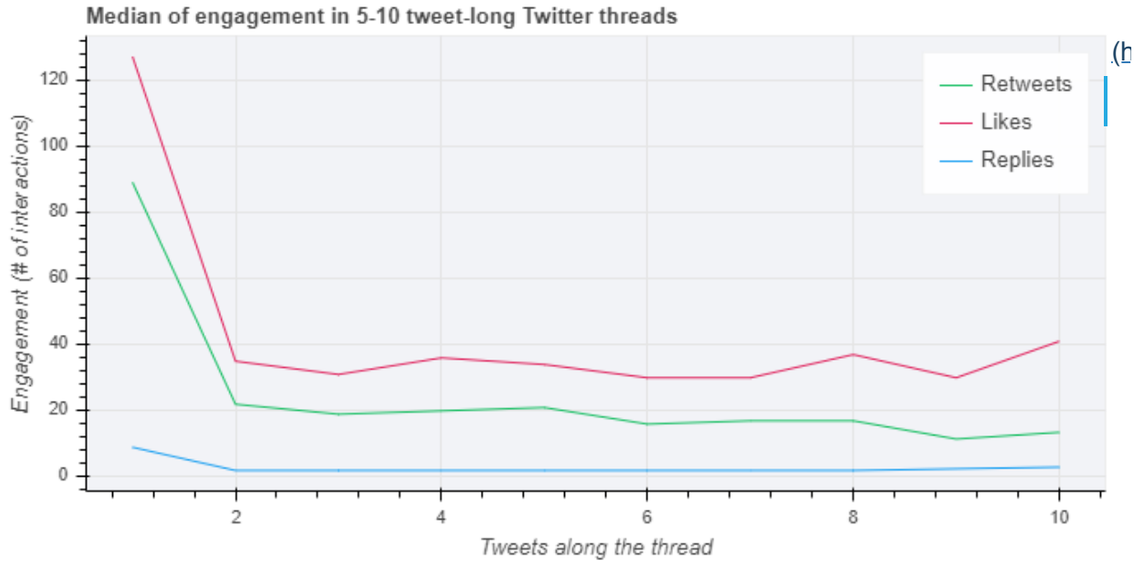
In [136]:

```
# median of engagement
median1 = pd.DataFrame()
median1['Retweets'] = retweets_by_thread1.median(axis=1)
median1['Likes'] = likes_by_thread1.median(axis=1)
median1['Replies'] = replies_by_thread1.median(axis=1)

median_engagement1 = figure(plot_width=700,
           plot_height=350,
           title='Median of engagement in 5-10 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

# add a line renderer
median_engagement1.line(list(range(1,11)), median1['Retweets'].values,line_color='#17bf6
3', legend='Retweets')
median_engagement1.line(list(range(1,11)), median1['Likes'].values, line_color='#e0245e'
, legend='Likes')
median_engagement1.line(list(range(1,11)), median1['Replies'].values, line_color='#1da1f
2', legend='Replies')

show(median_engagement1)
```



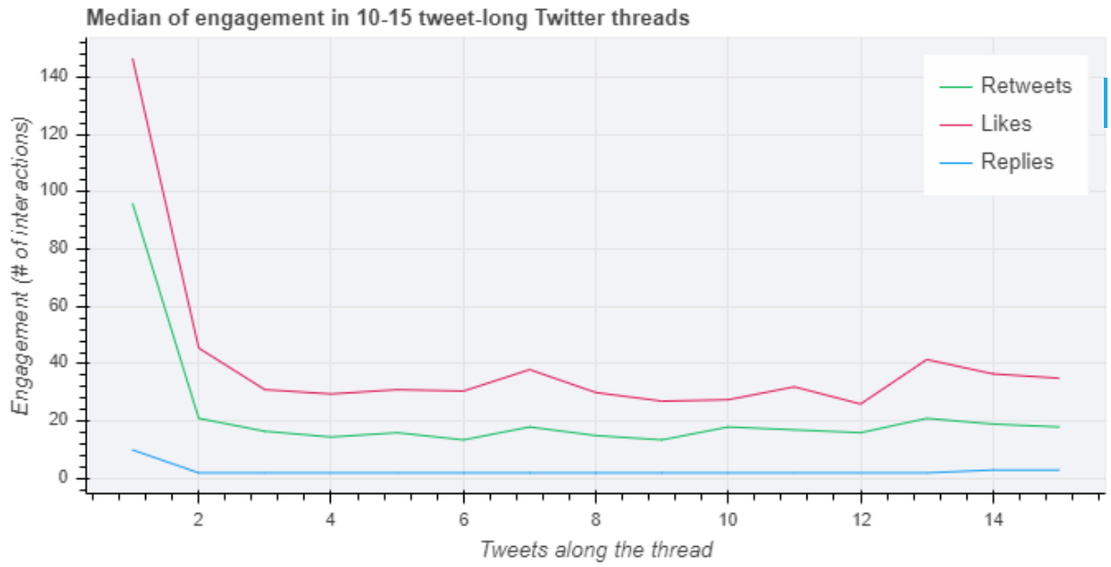
In [53]:

```
median2 = pd.DataFrame()
median2['Retweets'] = retweets_by_thread2.median(axis=1)
median2['Likes'] = likes_by_thread2.median(axis=1)
median2['Replies'] = replies_by_thread2.median(axis=1)

median_engagement2 = figure(plot_width=700,
           plot_height=350,
           title='Median of engagement in 10-15 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

# add a line renderer
median_engagement2.line(list(range(1,16)), median2['Retweets'].values,line_color='#17bf6
3', legend='Retweets')
median_engagement2.line(list(range(1,16)), median2['Likes'].values, line_color='#e0245e'
, legend='Likes')
median_engagement2.line(list(range(1,16)), median2['Replies'].values, line_color='#1da1f
2', legend='Replies')

show(median_engagement2)
```



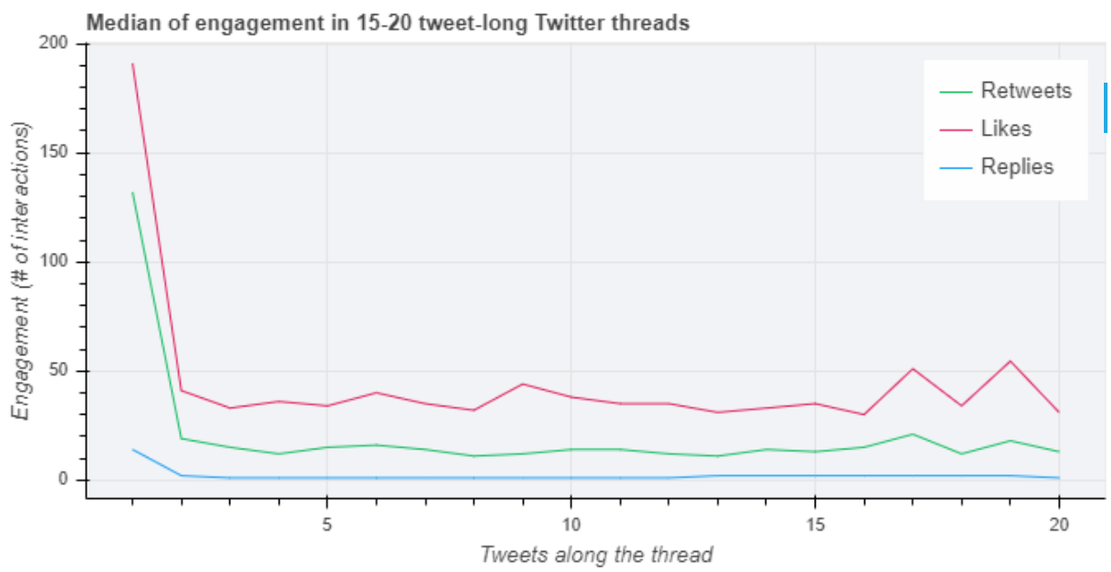
In [54]:

```
median3 = pd.DataFrame()
median3['Retweets'] = retweets_by_thread3.median(axis=1)
median3['Likes'] = likes_by_thread3.median(axis=1)
median3['Replies'] = replies_by_thread3.median(axis=1)

median_engagement3 = figure(plot_width=700,
           plot_height=350,
           title='Median of engagement in 15-20 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

# add a line renderer
median_engagement3.line(list(range(1,21)), median3['Retweets'].values,line_color='#17bf6
3', legend='Retweets')
median_engagement3.line(list(range(1,21)), median3['Likes'].values, line_color='#e0245e'
, legend='Likes')
median_engagement3.line(list(range(1,21)), median3['Replies'].values, line_color='#1da1f
2', legend='Replies')

show(median_engagement3)
```



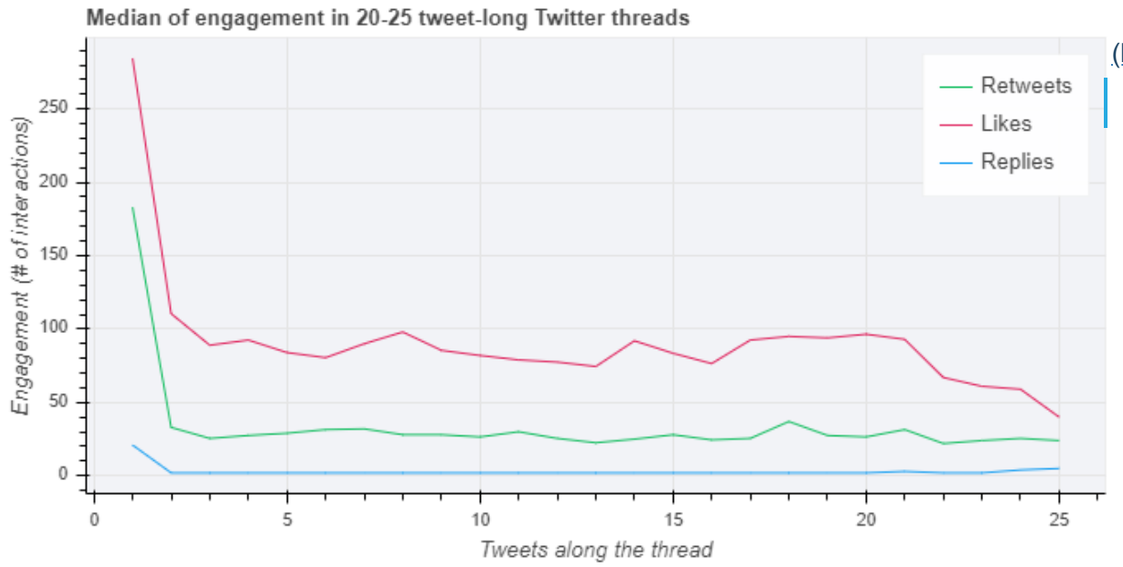
In [55]:

```
median4 = pd.DataFrame()
median4['Retweets'] = retweets_by_thread4.median(axis=1)
median4['Likes'] = likes_by_thread4.median(axis=1)
median4['Replies'] = replies_by_thread4.median(axis=1)

median_engagement4 = figure(plot_width=700,
           plot_height=350,
           title='Median of engagement in 20-25 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

# add a line renderer
median_engagement4.line(list(range(1,26)), median4['Retweets'].values,line_color='#17bf6
3', legend='Retweets')
median_engagement4.line(list(range(1,26)), median4['Likes'].values, line_color='#e0245e'
, legend='Likes')
median_engagement4.line(list(range(1,26)), median4['Replies'].values, line_color='#1da1f
2', legend='Replies')

show(median_engagement4)
```

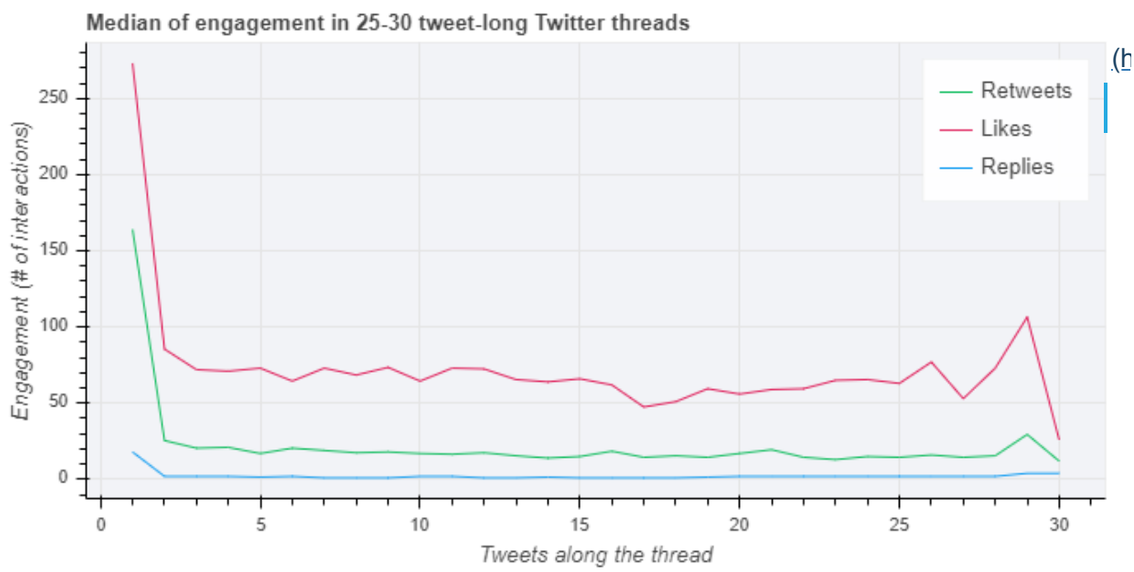


In [56]:
```
median5 = pd.DataFrame()
median5['Retweets'] = retweets_by_thread5.median(axis=1)
median5['Likes'] = likes_by_thread5.median(axis=1)
median5['Replies'] = replies_by_thread5.median(axis=1)

median_engagement5 = figure(plot_width=700,
           plot_height=350,
           title='Median of engagement in 25-30 tweet-long Twitter threads',
           background_fill_color="#f2f3f7",
           y_axis_label='Engagement (# of interactions)',
           x_axis_label='Tweets along the thread')

# add a line renderer
median_engagement5.line(list(range(1,31)), median5['Retweets'].values,line_color='#17bf6
3', legend='Retweets')
median_engagement5.line(list(range(1,31)), median5['Likes'].values, line_color='#e0245e'
, legend='Likes')
median_engagement5.line(list(range(1,31)), median5['Replies'].values, line_color='#1da1f
2', legend='Replies')

show(median_engagement5)
```



Conclusiones

Se observa que los likes y RT están relacionados, puesto que cuando los RT suben, los likes también. También se observa que el primer y los últimos tweets son los que mayor cantidad de reacciones poseen.

In [ ]: