# Gravity Model

Investigate W^{in}

W_b^{in} trend and normalization

- First we check a few landmarks and places with high level of tweets activities, observe their coefficients, i.e. W_b^{in} change during weekdays and weekends

- Second we select a place (here we simply choose the one with highest volume—10007, city hall) as our reference point and set its W_b^{in}" equal to 1, then re-run the OLS to find "new and normalized" W_b^{in} for each zip code b

- In the meantime we'll check if the sum of W_b^{in} changes over time or remain stable (first un-normalized case, then normalized one)

Stephen QIAN

The sum of W_b^{in} (un-normalized)

$\sum_b W_b^{in}(t)$ seems to be stable (around 50) if the data volume is high enough (>13000 entries, roughly speaking)
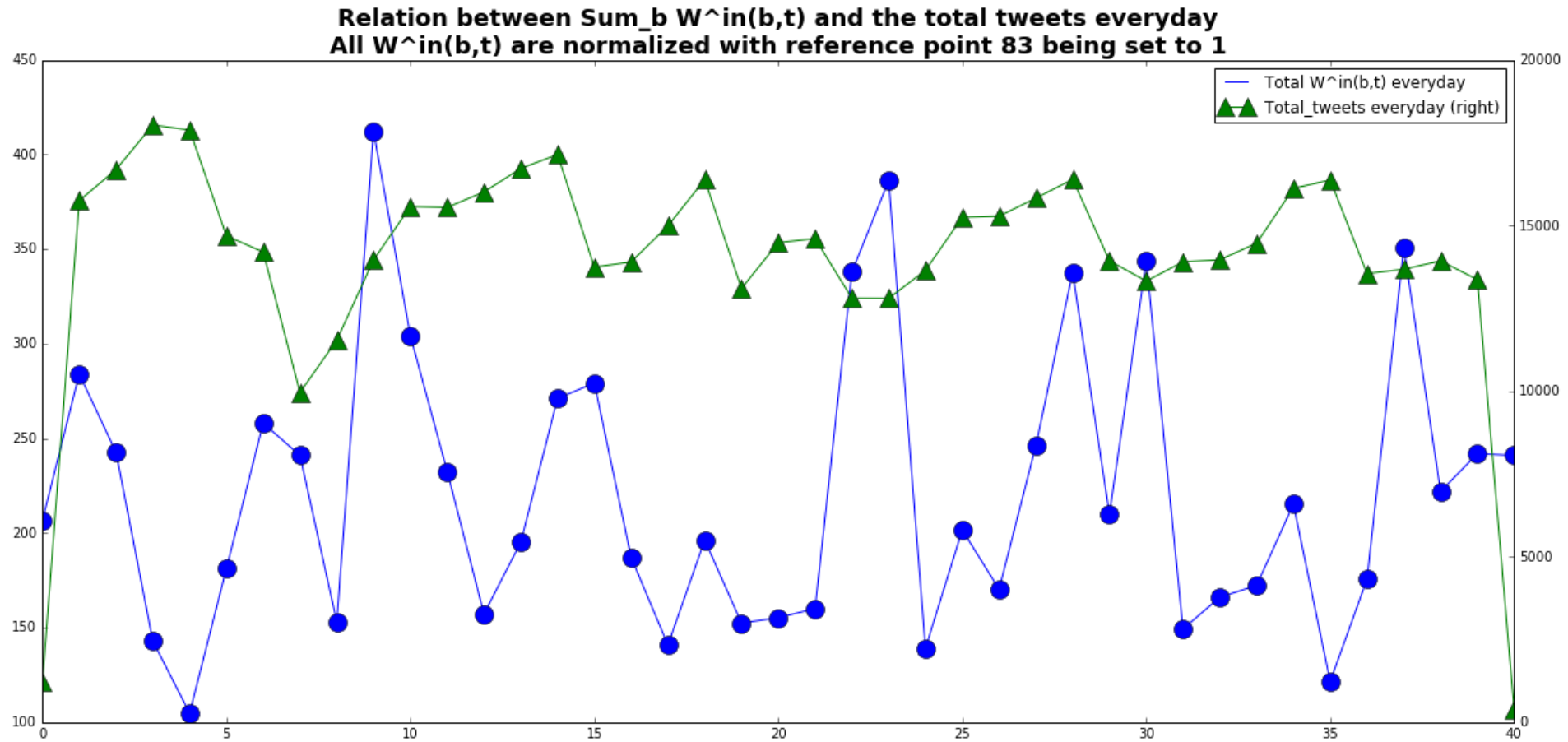

Relation between Sum_b W^in(b,t) and the total tweets everyday
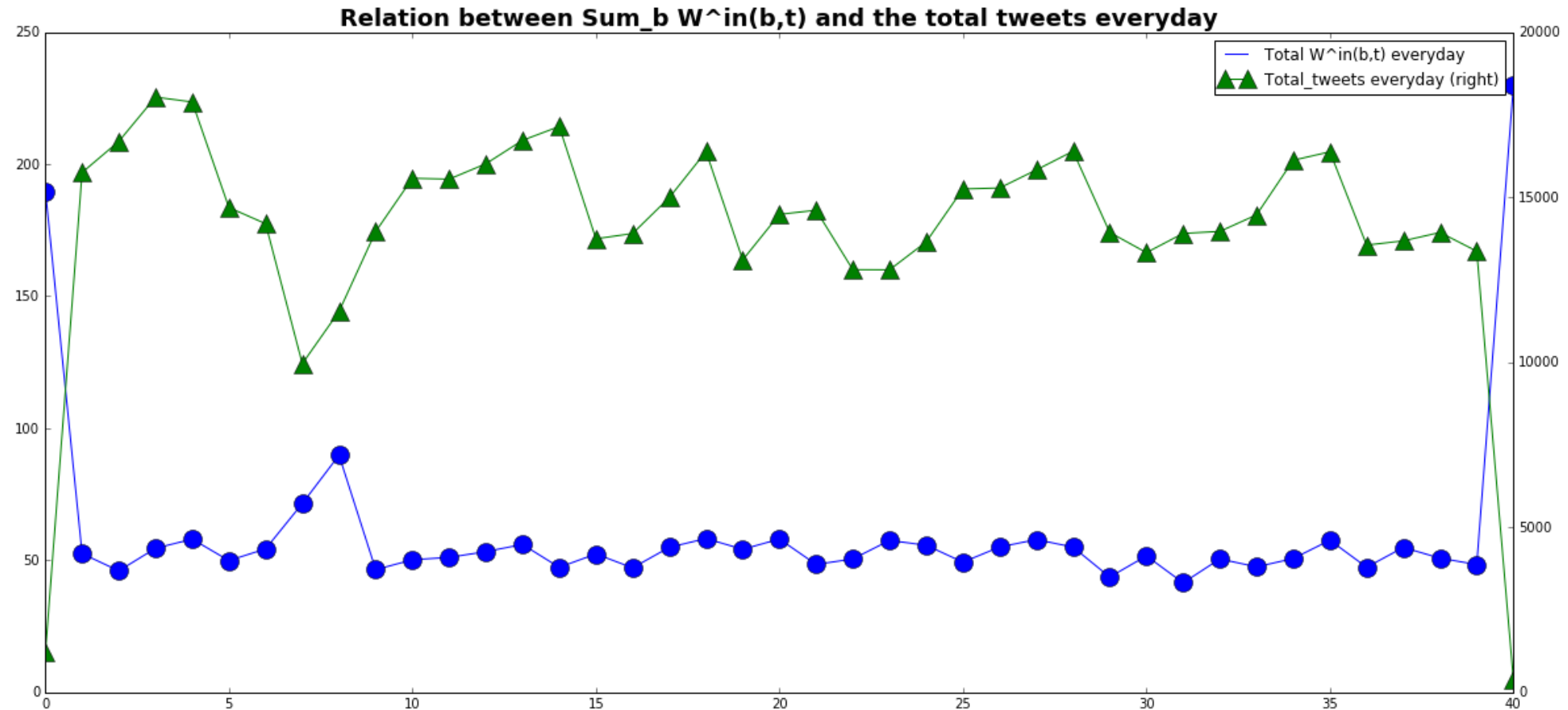
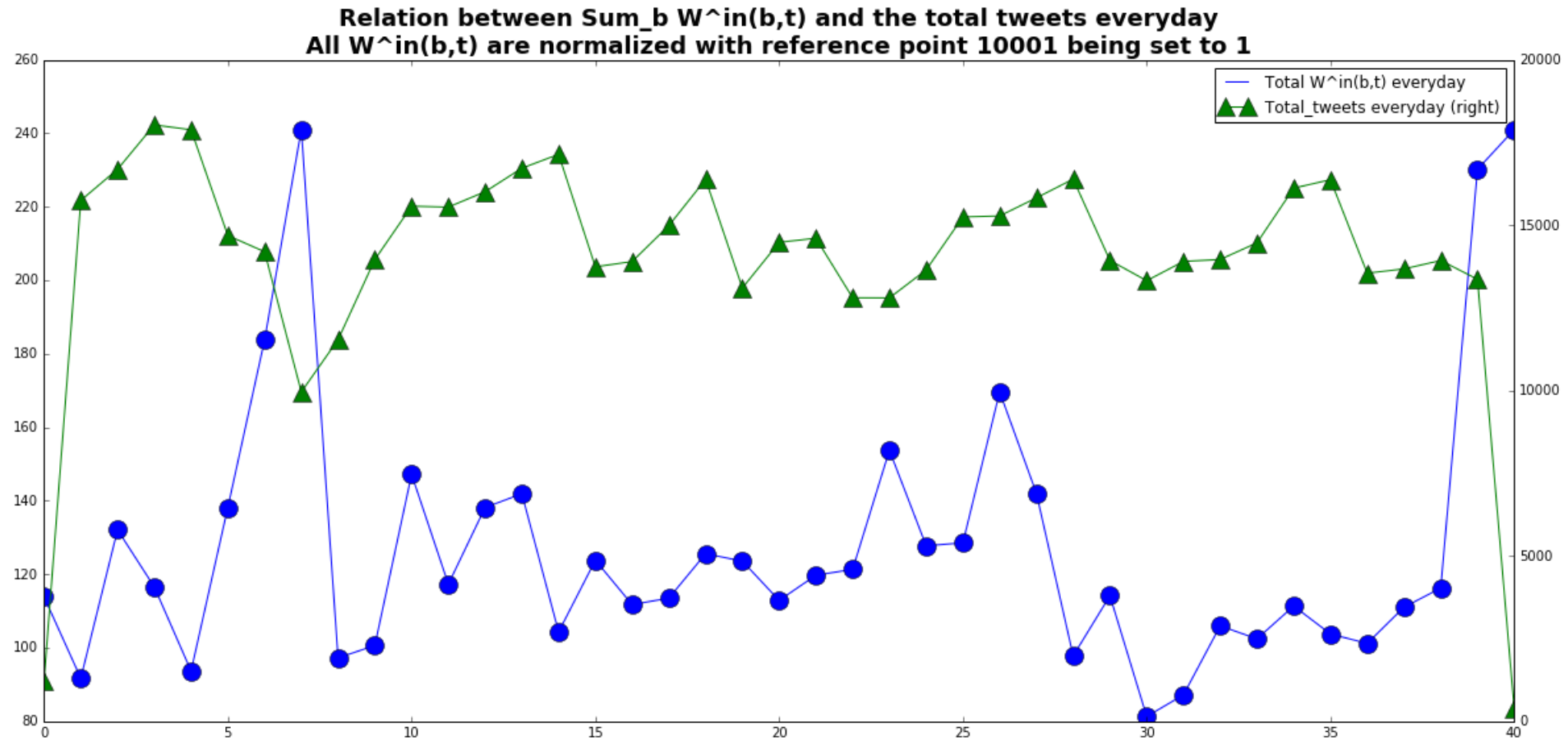# The sum of W_b^{in} (Normalization method A)

- Let each "new" W_b^{in} be $W_b^{in} / \sum_a W_a^{in}$
- Then of course now the sum would be 1 everyday (hence the plot is omitted)
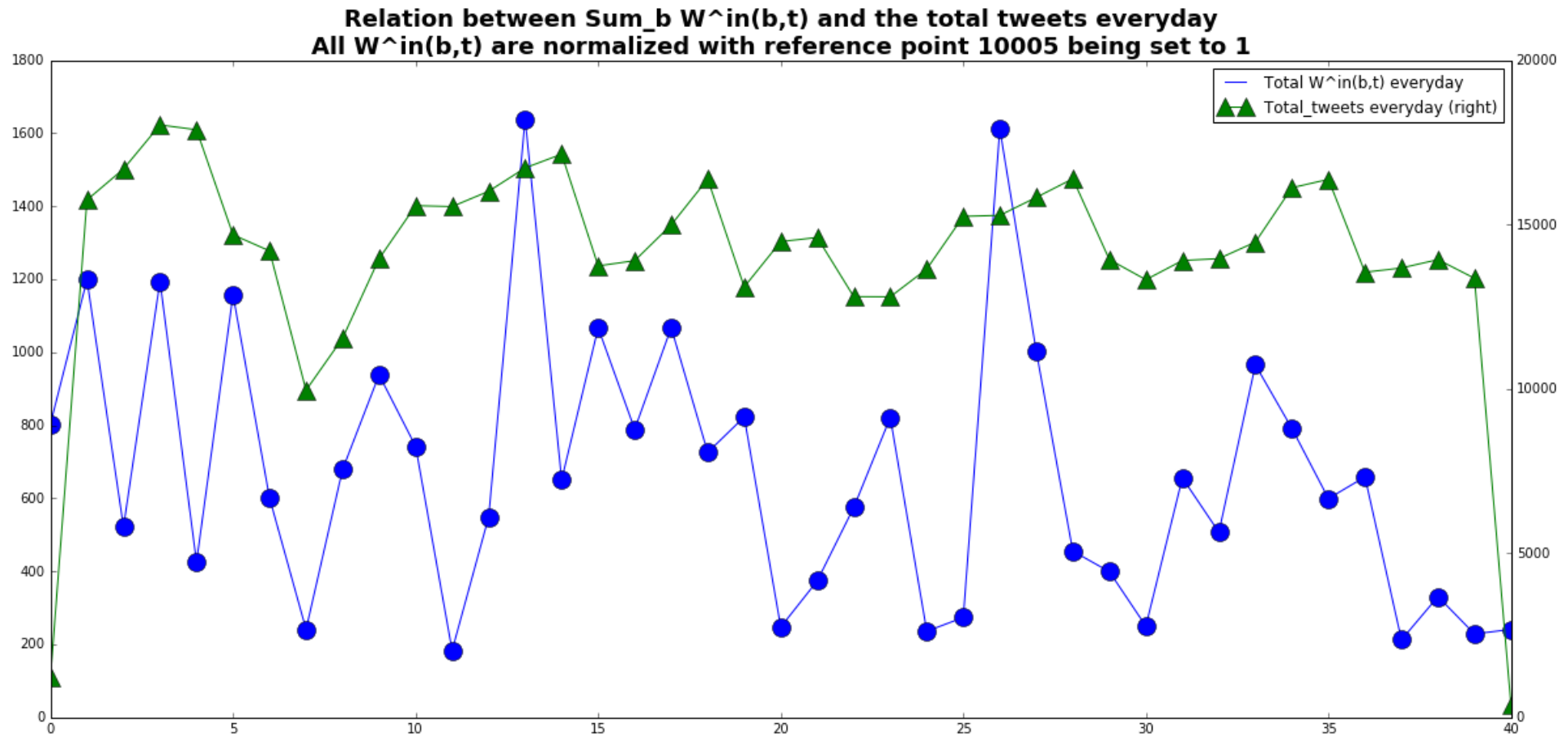
The sum of W_b^{in} (Normalization method B)

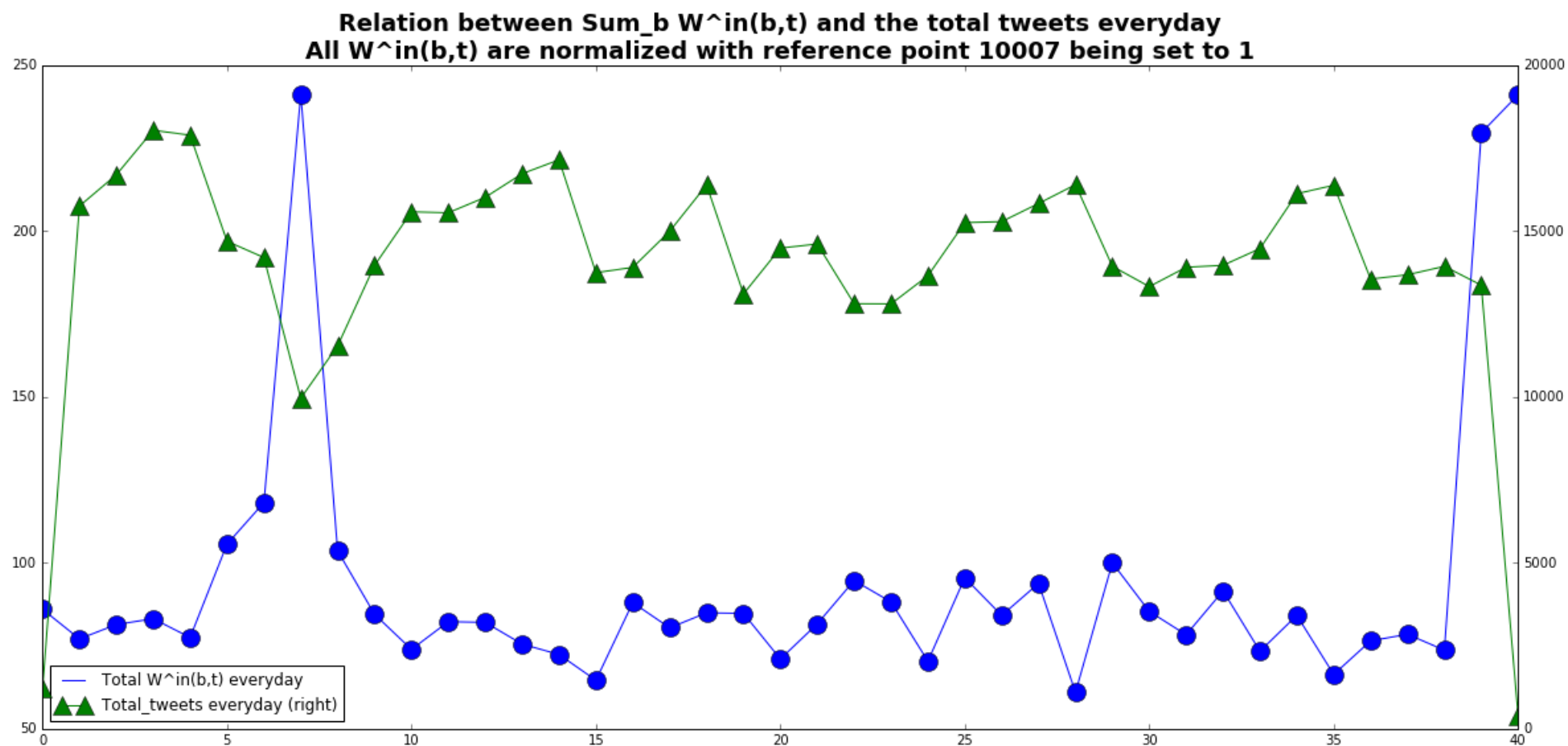Select a reference point/zip code, set its W_b^{in} equal to 1 then rerun the OLS to fit others

Relation between Sum_b W^in(b,t) and the total tweets everyday
All W^in(b,t) are normalized with reference point 83 being set to 1

(Normalization method B continued)

Relation between Sum_b W^in(b,t) and the total tweets everyday
All W^in(b,t) are normalized with reference point 10001 being set to 1

(Normalization method B continued)

Relation between Sum_b W^in(b,t) and the total tweets everyday
All W^in(b,t) are normalized with reference point 10005 being set to 1

(Normalization method B continued)

Relation between Sum_b W^in(b,t) and the total tweets everyday
All W^in(b,t) are normalized with reference point 10007 being set to 1

Legend: Total W^in(b,t) everyday; Total_tweets everyday (right)

(Normalization method B continued)

Relation between Sum_b W^in(b,t) and the total tweets everyday
All W^in(b,t) are normalized with reference point 10036 being set to 1

(Normalization method B continued)

Relation between Sum_b W^in(b,t) and the total tweets everyday
All W^in(b,t) are normalized with reference point 11201 being set to 1

W_b^{in} Summary

- It appears that method B is highly dependent on the point we choose, the sum of W_b^{in} varies dramatically and appears chaotic and random(among six zip codes we've tried, only 10007 presented relatively stable results, while showing obvious relation to the number of tweets every day)

- On the other hand, un-normalized values perform quite well as long as the data volume is high enough to generate accurate fitting, making normalization method A my personal preferable one.

- At last, we use method A to normalize and showcase those zip codes behaviors.
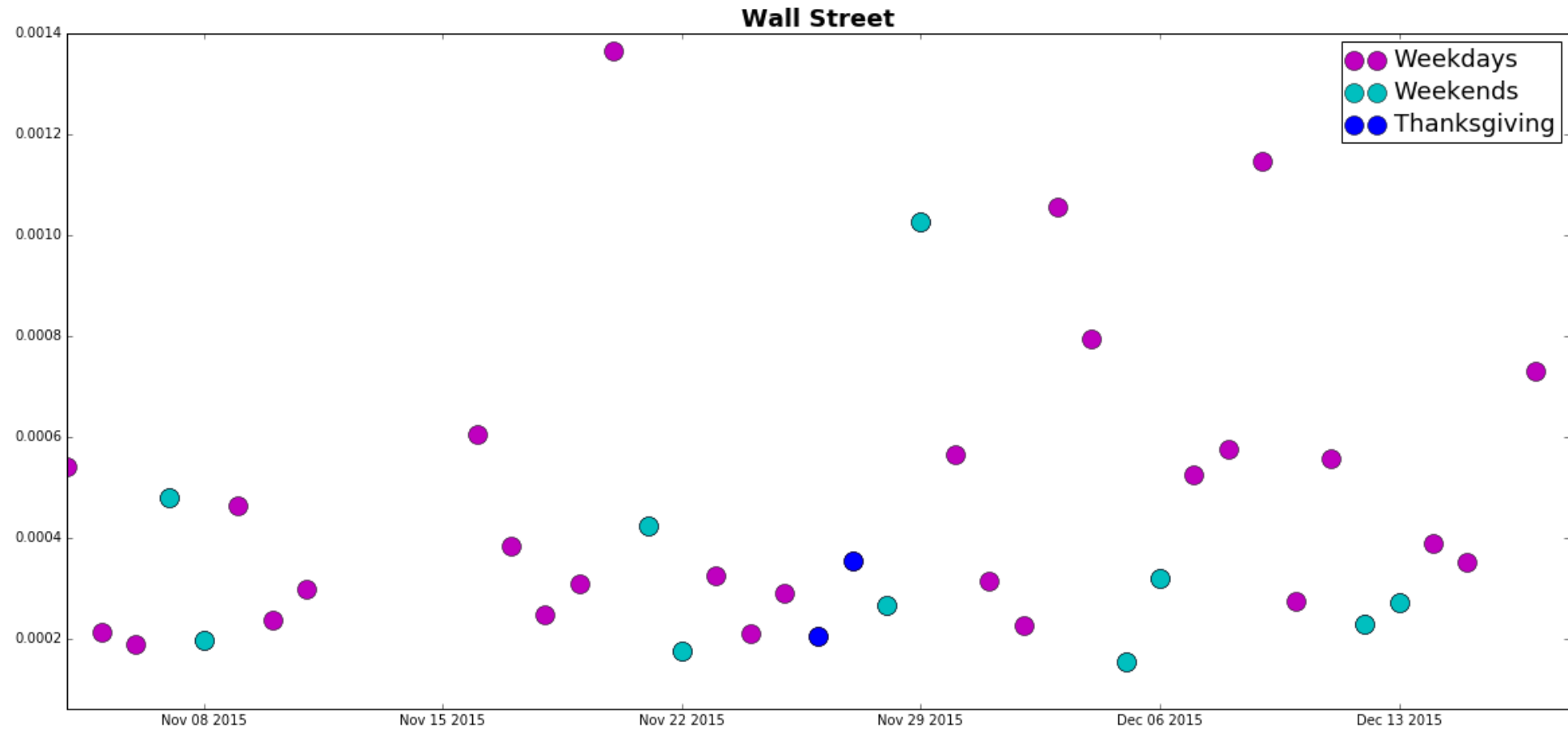
83, Central Park
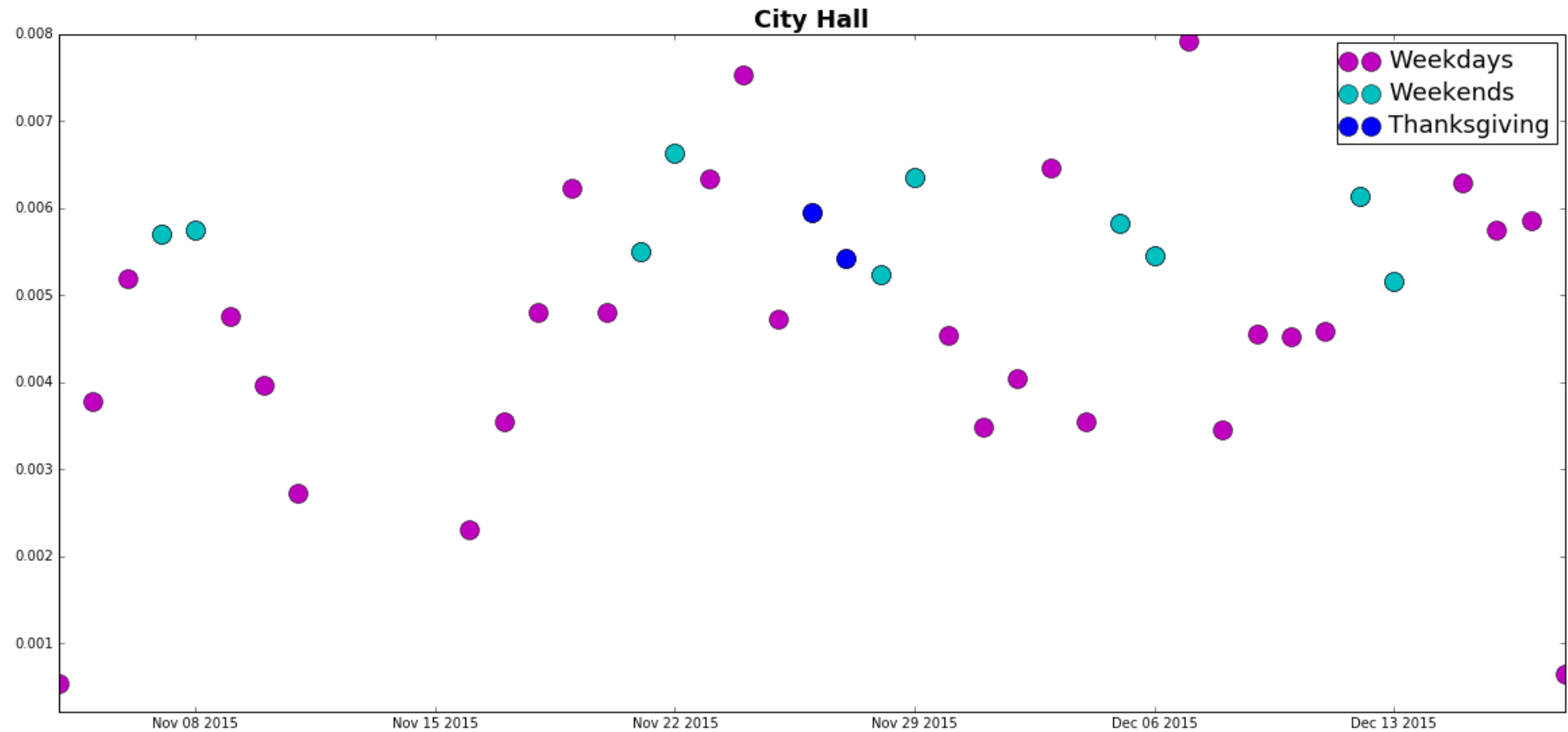# (Normalization method A, same for the rest)
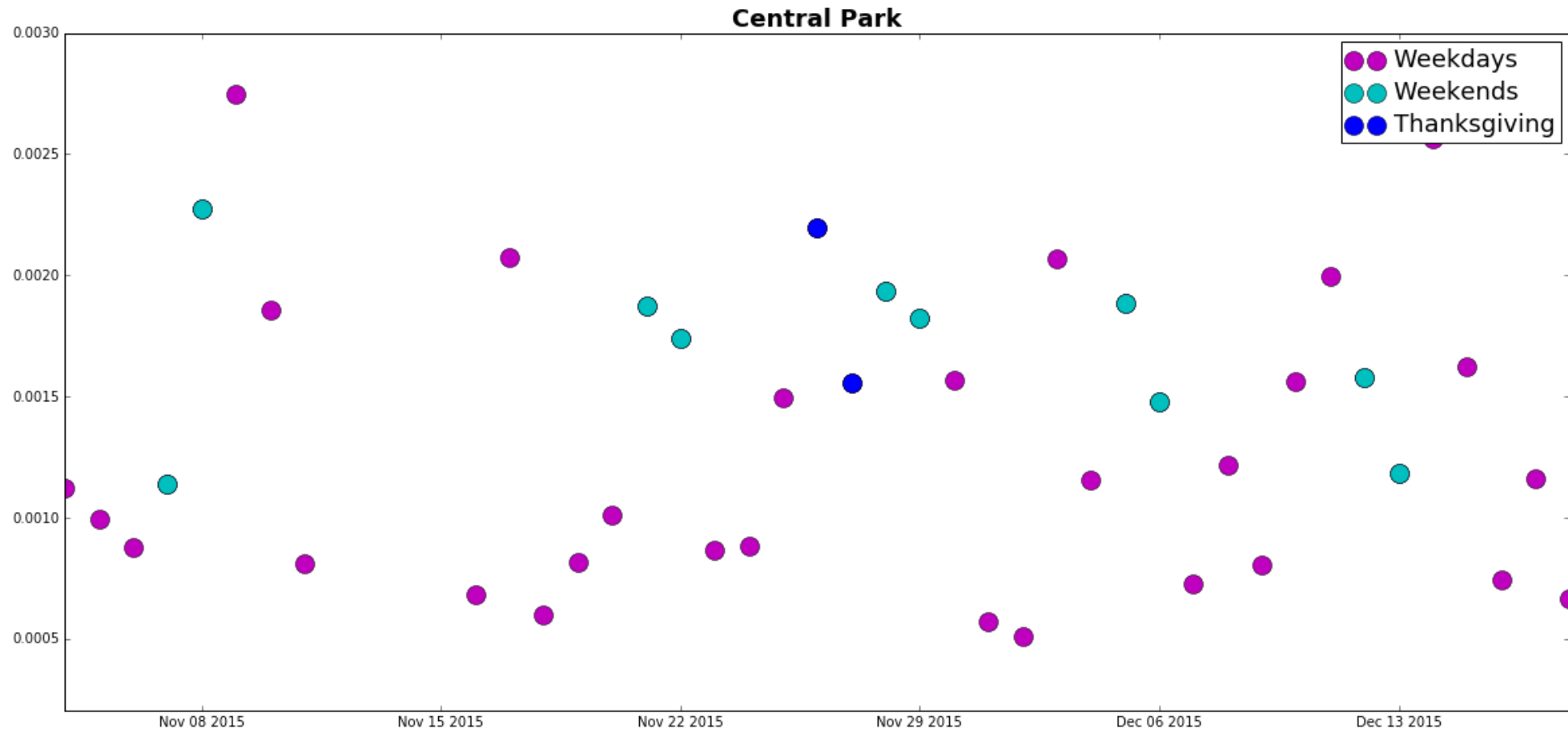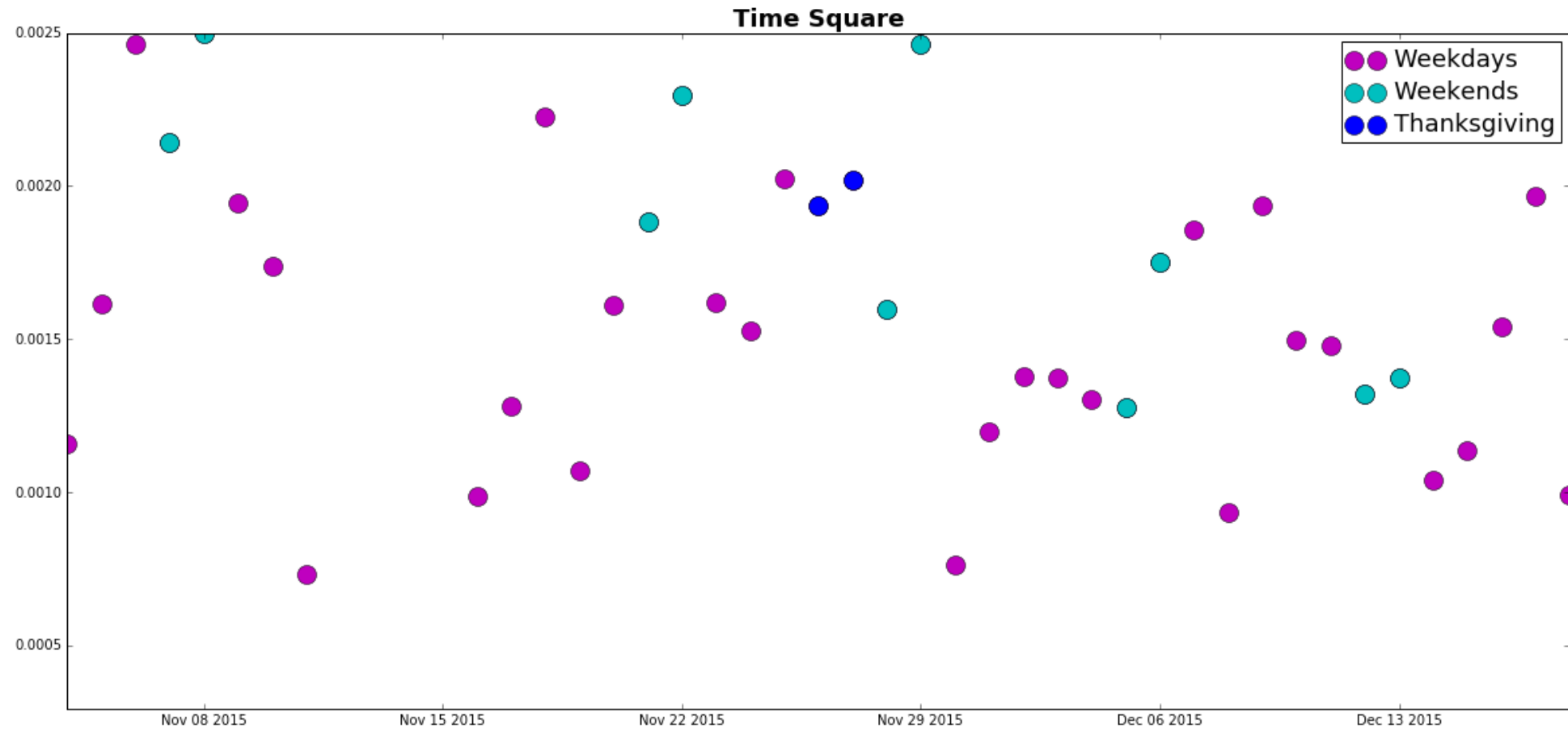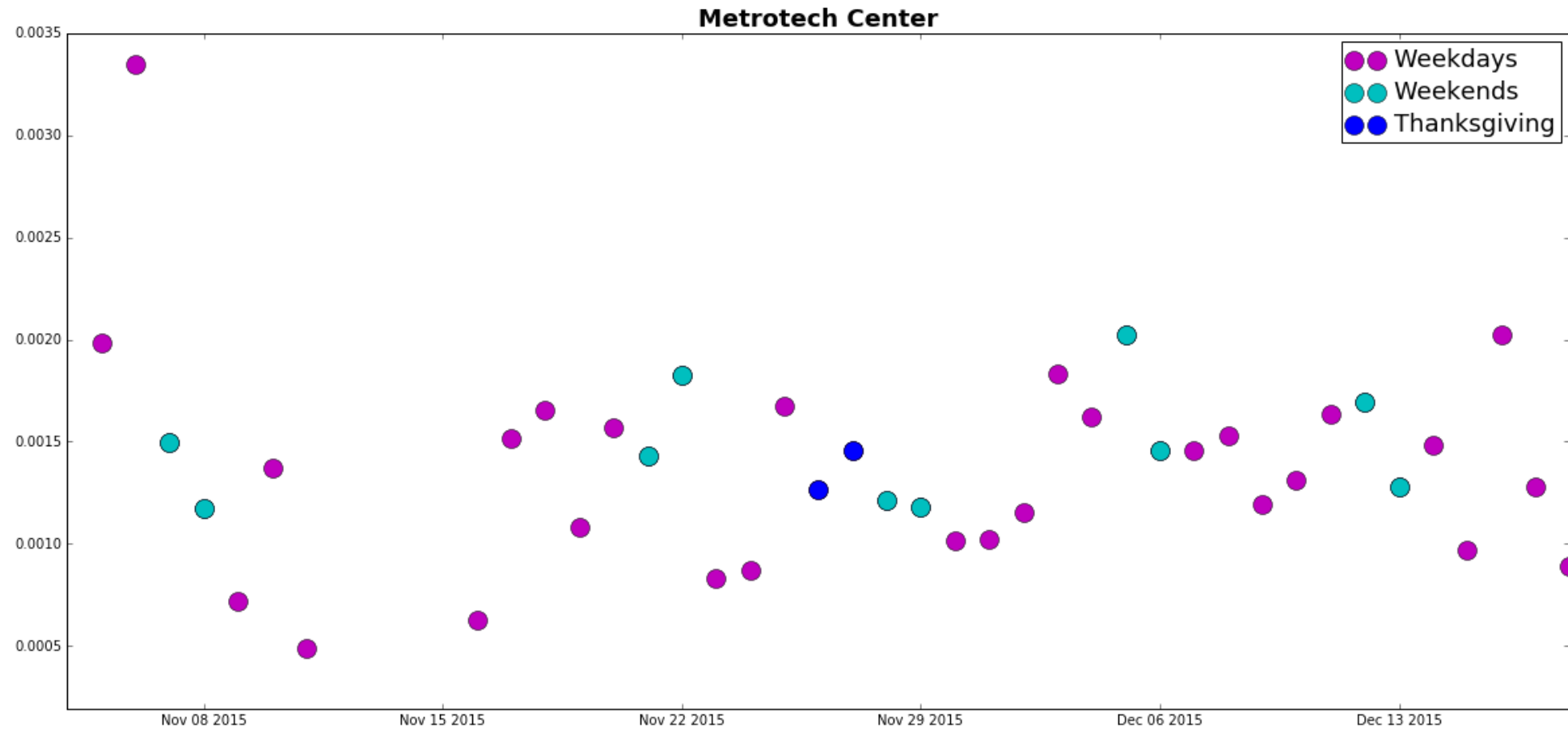
10001, Korean Town
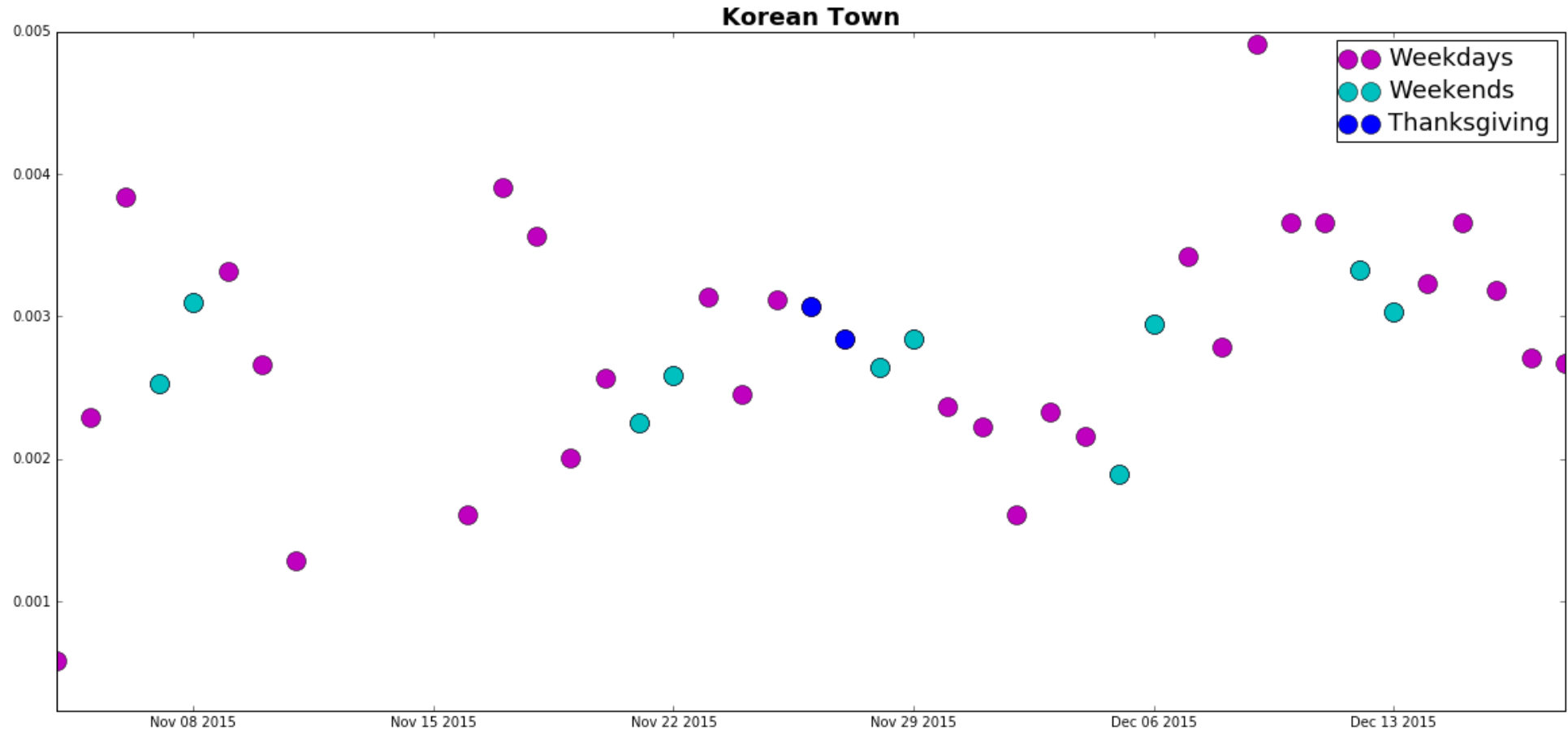
10005, Wall Street

10007, City Hall

10036, Time Square

11201, Metrotech Center

Appendix—The quantitative information (normalized)

	Weekdays mean	Weekdays std	Weekends mean	Weekends std	Thxgiving mean	Thxgiving std	location
83	0.001233	0.000610	0.001691	0.000350	0.001875	0.000451	Central Park
10001	0.002790	0.000907	0.002714	0.000426	0.002955	0.000161	Korean Town
10005	0.000500	0.000345	0.000354	0.000258	0.000280	0.000105	Wall Street
10007	0.004622	0.001783	0.005770	0.000481	0.005678	0.000370	City Hall
10036	0.001462	0.000437	0.001860	0.000471	0.001977	0.000057	Time Square
11201	0.001316	0.000604	0.001478	0.000291	0.001364	0.000135	Metrotech Center

Appendix—Summary Visualization

Quantitative Summary of normalized W^in in some selected zip codes