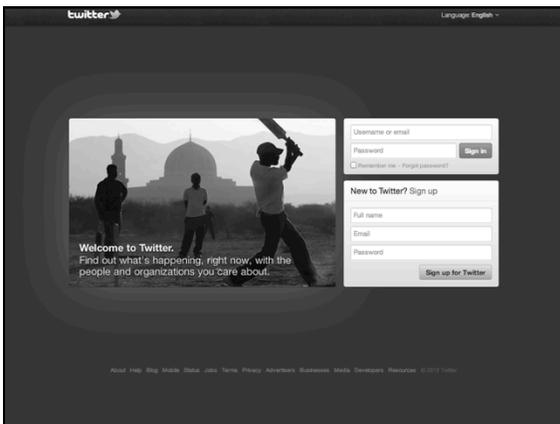
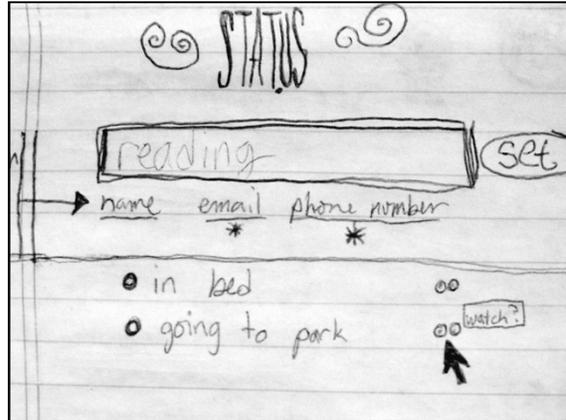


instantly connect
people everywhere
 to what's most meaningful to them



Magnitude	9.0
Date-Time	Friday, March 11, 2011 at 05:46:24 UTC Friday, March 11, 2011 at 02:46:24 PM at epicenter Time of Earthquake in other Time Zones
Location	38.297°N, 142.372°E
Depth	30 km (18.6 miles) set by location program
Region	NEAR THE EAST COAST OF HONSHU, JAPAN
Distances	129 km (80 miles) E of Sendai, Honshu, Japan 177 km (109 miles) E of Yamagata, Honshu, Japan 177 km (109 miles) ENE of Fukushima, Honshu, Japan 373 km (231 miles) NE of TOKYO, Japan
Location Uncertainty	horizontal +/- 13.5 km (8.4 miles); depth fixed by location program
Parameters	NST=350, Nph=351, Dmin=416.3 km, Rmss=0.1 sec, Cp= 29°, M-type=centroid moment magnitude (Mw), Version=A
Source	USGS NEIC (WDCS-D)
Event ID	usc0001xgp

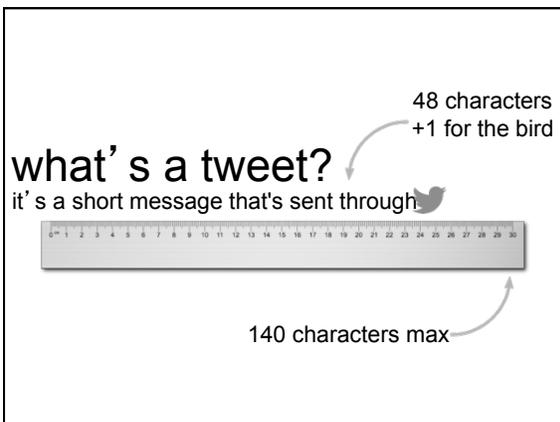




Tweets have a max length of

A.
 B.
 C.

characters are not the same as bytes!

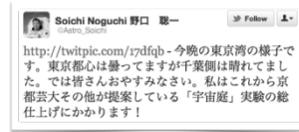


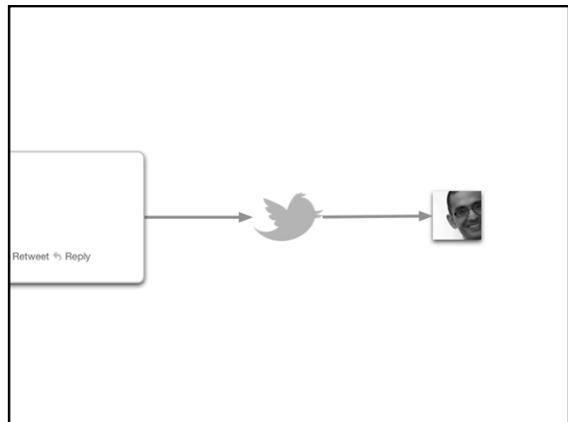
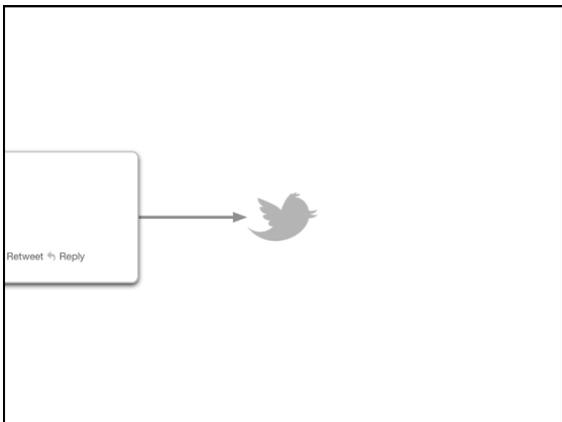
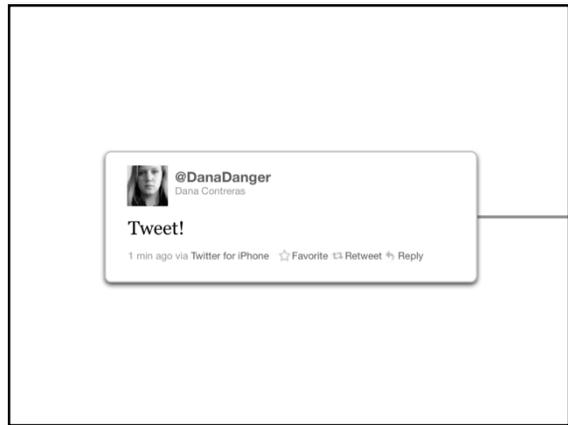
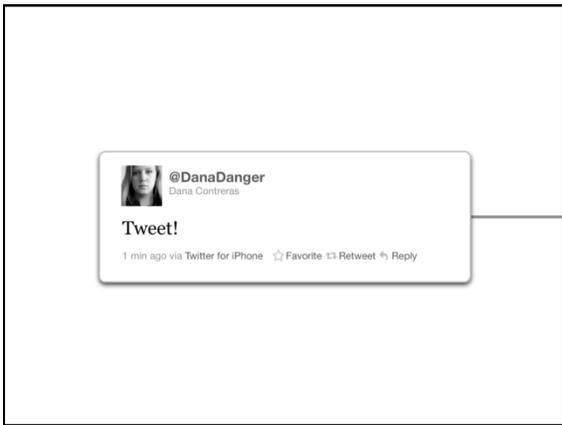
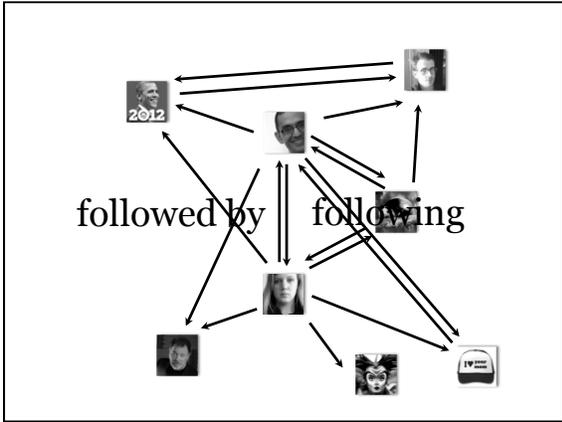
1 tweet

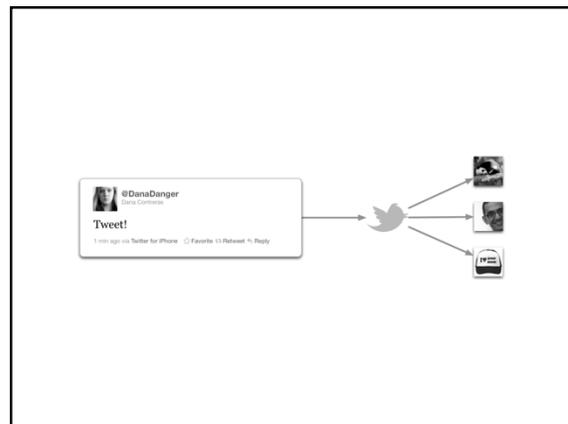
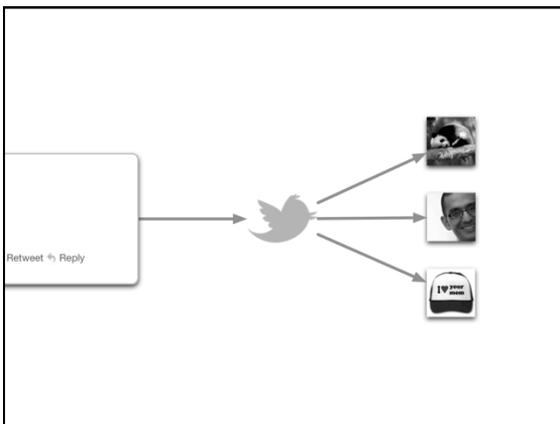
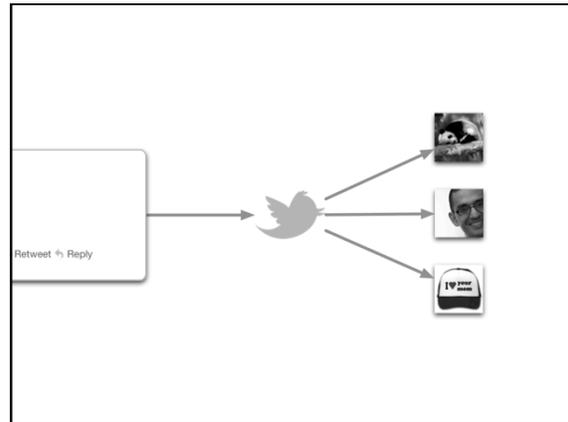
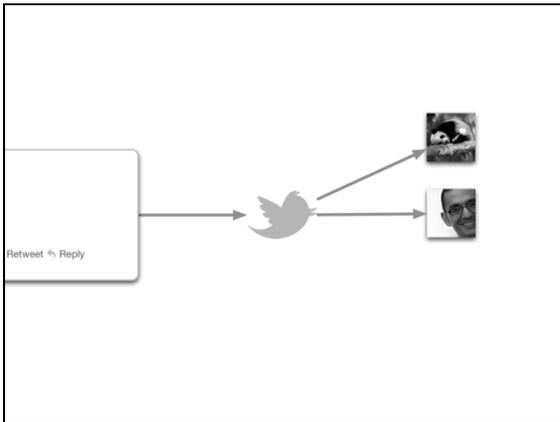
sizeof (1 tweet)

sizeof (1 tweet) \approx 200 bytes

sizeof (1 tweet) = 140 characters
 \approx 200 bytes







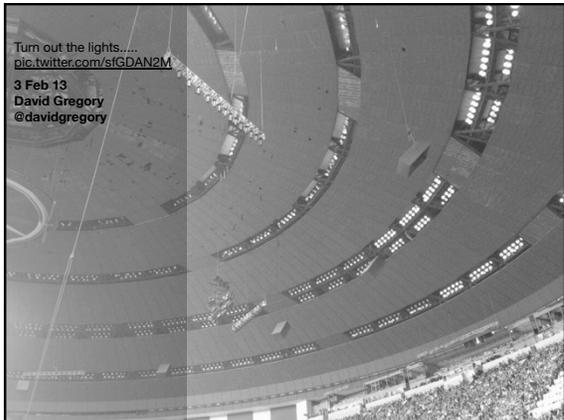
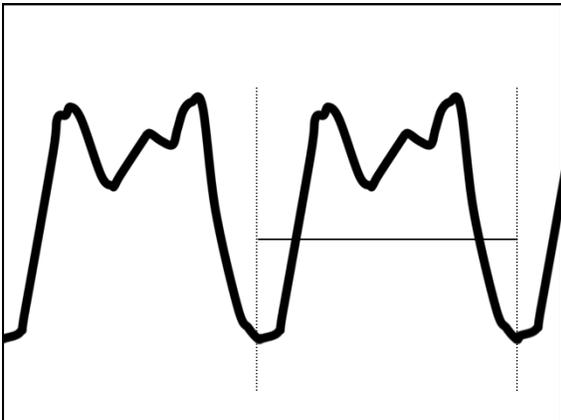
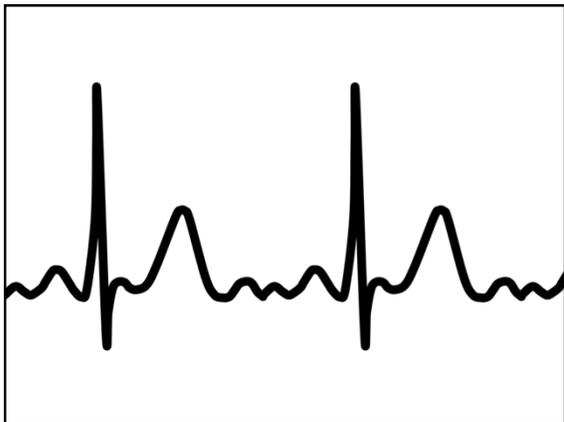
sounds easy, right? 
trick question

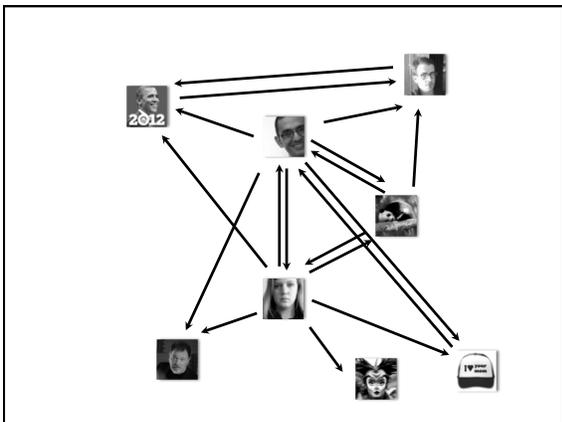
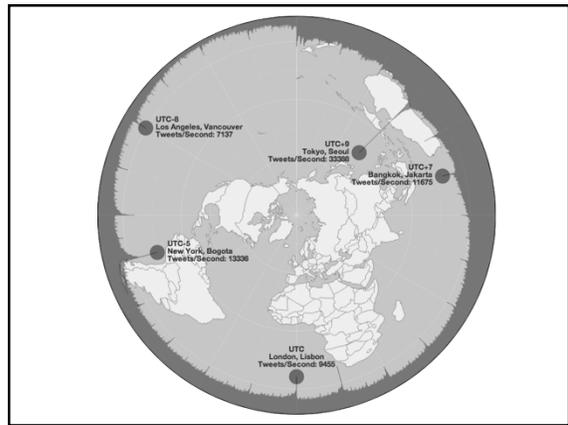
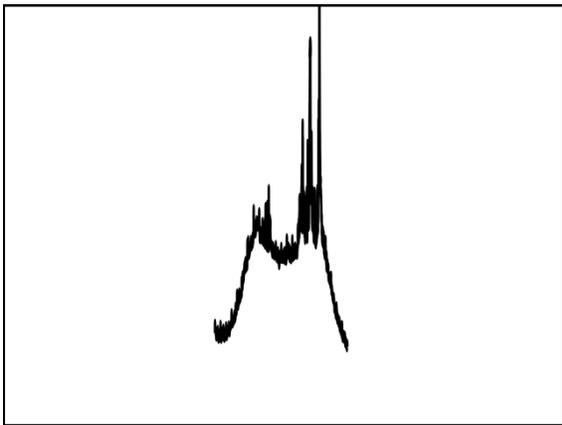
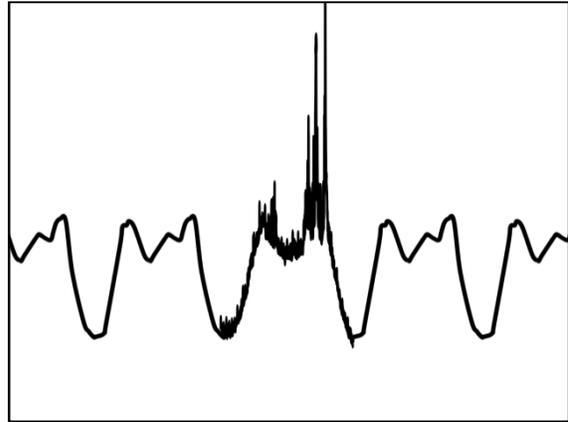
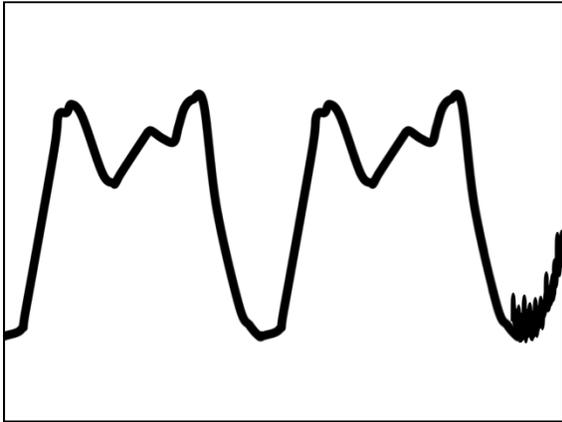
How many tweets are there?
 A. 50 a second
 B. 500 a minute
 C. 5 million an hour
 D. 500 million a day

about
500,000,000
tweets a day

$5.0E6$ tweets / day \approx 5500 tweets / second
steady state \curvearrowright

my heart beats
90,000
times a day





What's the max vertex degree?

- A. 0
- B. 10
- C. 1 million
- D. There is no max

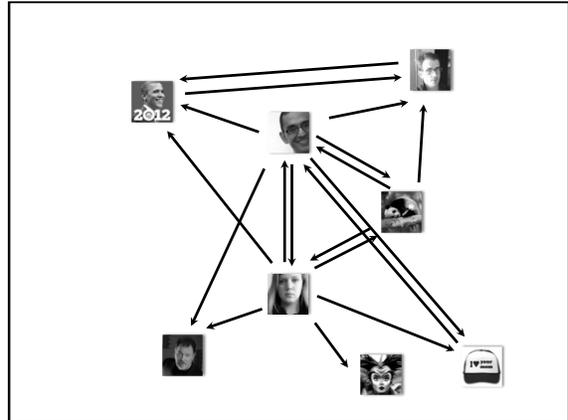


 **@justinbieber**
36 million followers

 **@ladygaga**
35 million followers

 **@BarackObama**
28 million followers

 **@riotun**
0.000002 million followers



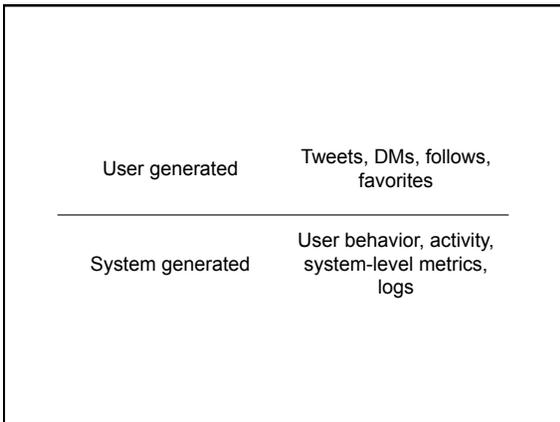
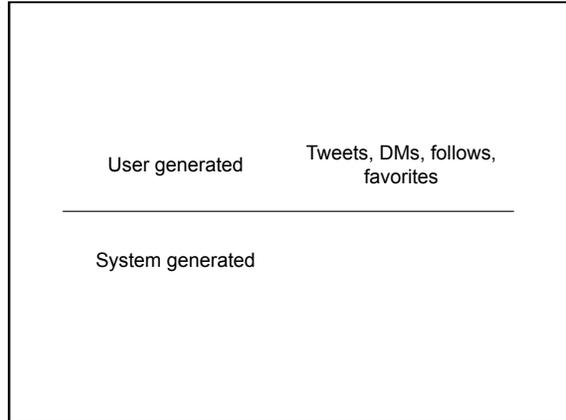
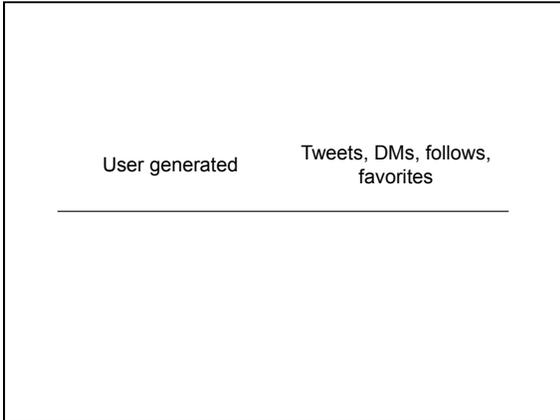
about
BILLIONS
of deliveries a day

We have a lot of
data



User generated

User generated Tweets, DMs, follows,
favorites

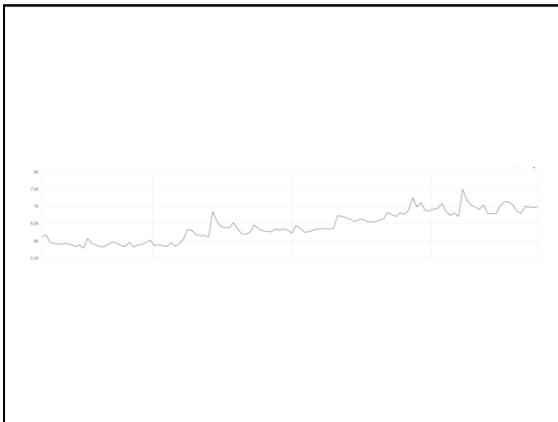
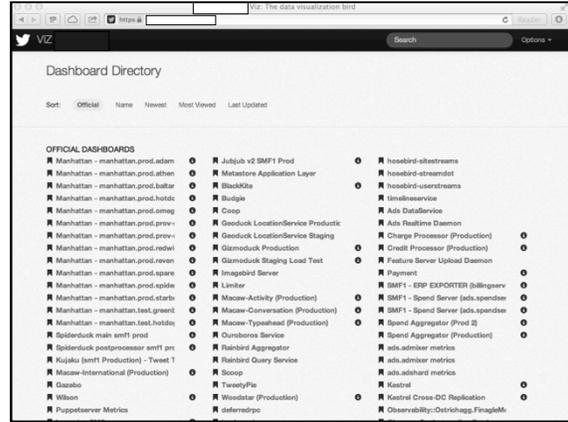


What's the ratio of user to system generated data?

A. 10:1
B. 1:1
C. 1:50
D. 1:1000

What are the questions we want to answer?





Batch vs Real-time

Batch: “A quantity or number taken together”

- think search engine: building the search index
- ‘solved problem’: Google’s MapReduce

Real-time: “Now. This instant”

- knowing what to look for in advance
- Twitter is a specific implementation of a “real-time” system

what happens in the
world
 happens on

Thank you!

Questions?

@rlotun