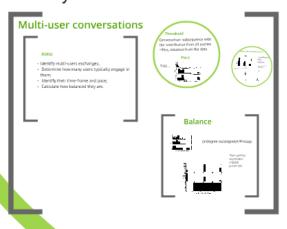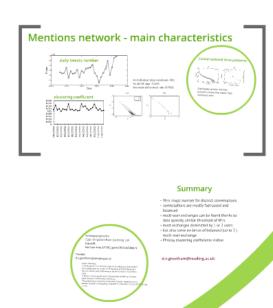# On Twitter conversations

Danica Greetham, University of Reading

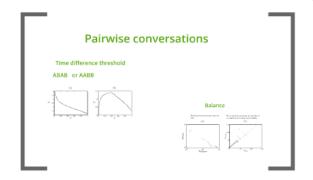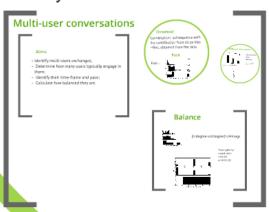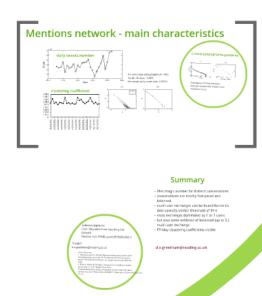joint work with Jonathan Ward,
University of Leeds

d.v.greetham@reading.ac.uk

Motivation

"everything is about conversation and not about broadcasting"

Most of the previous work about information flow and retweeting...
Most of the previous network analysis focused on followers...

"A visit from the Goon Squad" by Jennifer Egan, 2011

Aim
To describe main characteristics of Twitter "chatter": pace, timing, balance

data

We looked at the tweets from UK to UK having @ sign

If person A puts "@B" it designates that A is addressing the tweet to B specifically, but B is not the only one who can see the message (even people who do not follow A could see this message).

numbers:
4.4M, 730k,
daily:
vertices: 66877 --- 109213 (87624),
edges: 133308 --- 300295 (158380)

The data was collected by Datasift.

Pairwise conversations

Time difference threshold

ABAB or AABB

Balance

Multi-user conversations

Aims:
- Identify multi-users exchanges;
- Determine how many users typically engage in them;
- Identify their time-frame and pace;
- Calculate how balanced they are.

Threshold

Conversation-subsequence with the contribution from all parties ~9hrs, obtained from the data

Pace

Balance

(indegree-outdegree)/#messags

Mentions network - main characteristics

daily tweets number

clustering coefficient

Conversational time-patterns

Summary

- 9hrs magic number for distinct conversations
- conversations are mostly fast-paced and balanced
- multi-user exchanges can be found thanks to data sparsity, similar threshold of 9hrs
- most exchanges dominated by 1 or 2 users
- but also some evidence of balanced (up to 5) multi-user exchange
- FFriday clustering coefficients visible

d.v.greetham@reading.ac.uk

On Twitter conversations

Danica Greetham, University of Reading

joint work with Jonathan Ward,
University of Leeds

d.v.greetham@reading.ac.uk

Motivation

"everything is about conversation and not about broadcasting"

Most of the previous work about information flow and networking...
Most of the previous network analysis focused on followers...

"A visit from the Goon Squad" by Jennifer Egan, 2011

Aims:
to describe main characteristics of Twitter "chatter": pace, timing, balance

data

We looked at the tweets from UK to UK having @ sign

If person A puts "@B" it designates that A is addressing the tweet to B specifically, but B is not the only one who can see the message (even people who do not follow A could see this message).

numbers:
4.4M, 710k;
daily:
vertices: 66877 — 109213 (87624);
edges: 123398 — 200295 (158590)

The data was collected by Datasift.

Pairwise conversations

Time difference threshold

ABAB or AABB

Balance

Multi-user conversations

Aims:
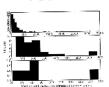- Identify multi-users exchanges;
- Determine how many users typically engage in them;
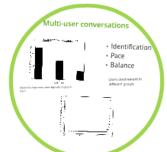- Identify their time-frame and pace;
- Calculate how balanced they are.

Threshold

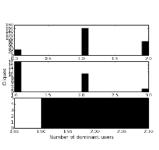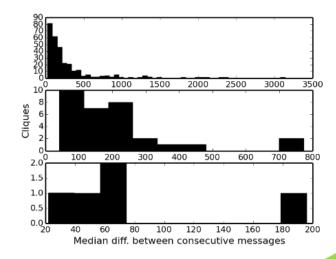Conversation: subsequence with the contribution from all parties =N+o, obtained from the data

Pace

Balance

|indegree-outdegree|+intrags

Mentions network - main characteristics

daily tweets number

Conversational time-patterns

clustering coefficient

Summary

- Min magic number for distinct conversations
- conversations are mostly fast-paced and balanced
- multi-user exchanges can be found thanks to data sparsity, similar threshold of 9mins
- most exchanges dominated by 1 or 2 users
- but also some evidence of balanced (up to 5) multi-user exchange
- FFriday clustering coefficients visible

d.v.greetham@reading.ac.uk

data

We looked at the tweets from UK to UK having @ sign

If person A puts "@B" it designates that A is addressing the tweet to B specifically, but B is not the only one who can see the message (even people who do not follow A could see this message).

numbers

4.4M, 730k,

daily:

vertices: 66877 --- 109213 (87624),

edges: 123398 --200295 (158380)

The data was collected by Datasift.

December 2011 ‹ ›

S	M	T	W	T	F	S
27	28	29	30	1	2	3
4	5	6	7	8	9	10
11	12	**13**	14	**15**	**16**	17
18	19	20	21	22	23	**24**
25	**26**	27	28	29	30	**31**
1	2	3	4	5	6	7

Motivation

``everything is about conversation and not about broadcasting''

Most of the previous work about information flow and retweeting...

Most of the previous network analysis focused on followers...

"A visit from the Goon Squad" by Jennifer Egan, 2011

"Those pauses, like the spaces between PowerPoint slides, become a metaphor for the gaps between what we mean and what we say"

Aim
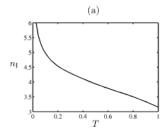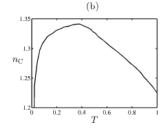
to describe main characteristics of Twitter "chatter": pace, timing, balance

Pairwise conversations

Time difference threshold

ABAB   or AABB

Balance

Time difference threshold

ABAB   or AABB

Balance

Distribution of conversation balance ratio

(a)
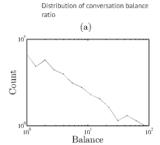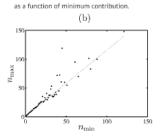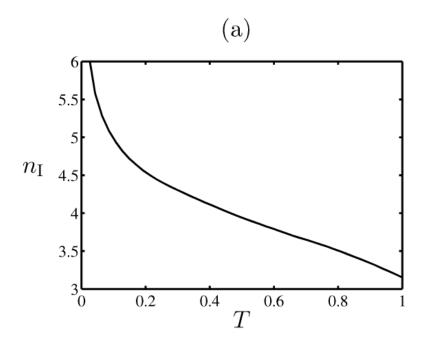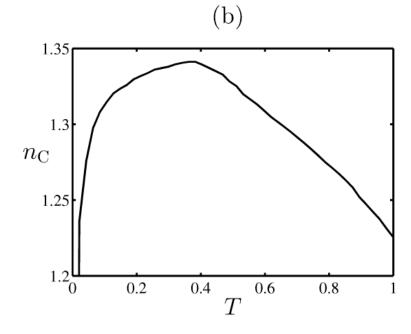
Mean maximum conversation contribution as a function of minimum contribution.

(b)

Multi-user conversations

Aims:

- Identify multi-users exchanges;
- Determine how many users typically engage in them;
- Identify their time-frame and pace;
- Calculate how balanced they are.

Threshold

Conversation- subsequence with the contribution from all parties ~9hrs, obtained from the data

Pace

Fast...

Multi-user conversations
- Identification
- Pace
- Balance

Balance

(indegree-outdegree)+/#mssgs

'floor-gaining'
round robin
r=|E|/|V|
p=100*r/|E|

Aims:

- Identify multi-users exchanges;
- Determine how many users typically engage in them;
- Identify their time-frame and pace;
- Calculate how balanced they are.

Threshold

Conversation- subsequence with the contribution from all parties ~9hrs, obtained from the data

Pace

Fast...

Multi-user conversations

No. of instances, sizes of cliques
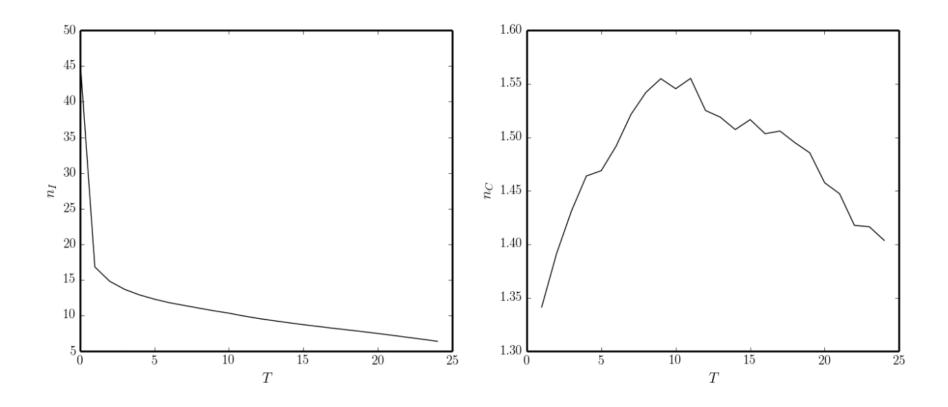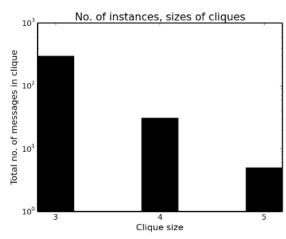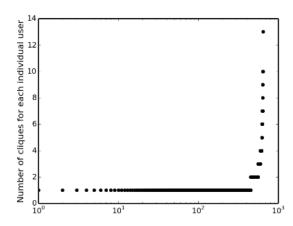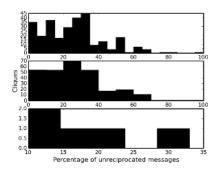
- Identification
- Pace
- Balance
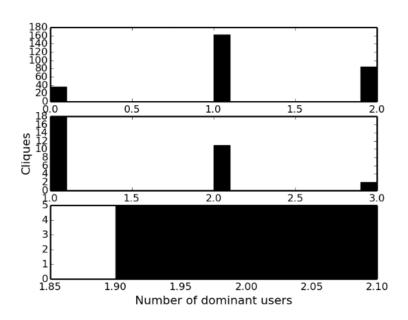
Users involvement in different groups

Determine how many users typically engage in them

Balance

(indegree-outdegree)+/#mssgs
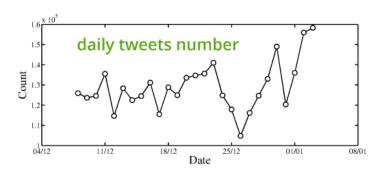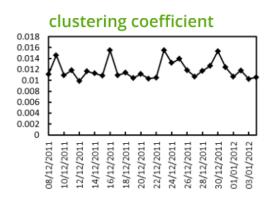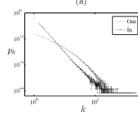
'floor-gaining'
round robin
r=|E|/|V|
p=100*r/|E|

Mentions network - main characteristics

daily tweets number

An individual daily maximum -590;
for all 28 days - 5,604;
the mean daily tweet rate -0.9410.

clustering coefficient

Conversational time-patterns

Distribution of time intervals
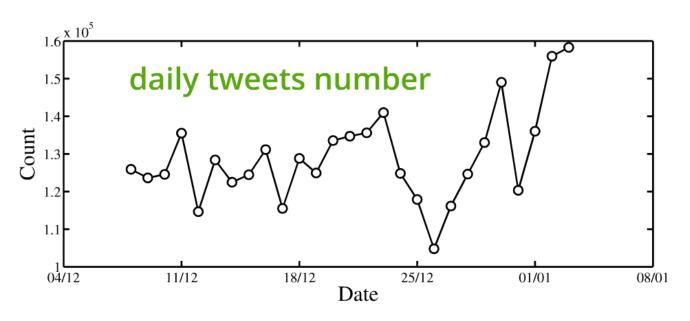between consecutive tweets from
individual users

daily tweets number

An individual daily maximum -590;
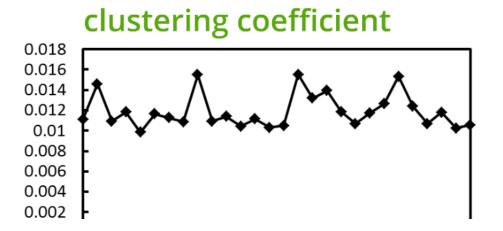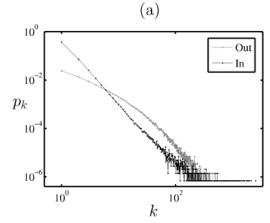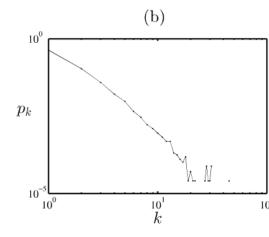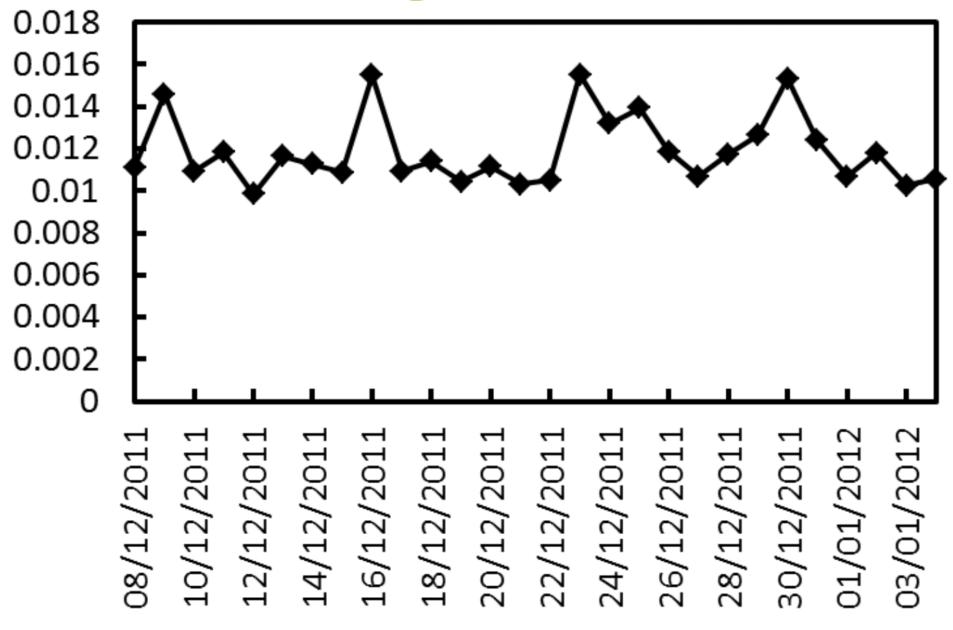for all 28 days - 5,604;
the mean daily tweet rate -0.9410.

clustering coefficient

(a) (b)

entions network - main char

daily tweets number

An individual daily maximum -590;
for all 28 days - 5,604;
the mean daily tweet rate -0.9410.

clustering coefficient
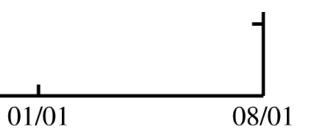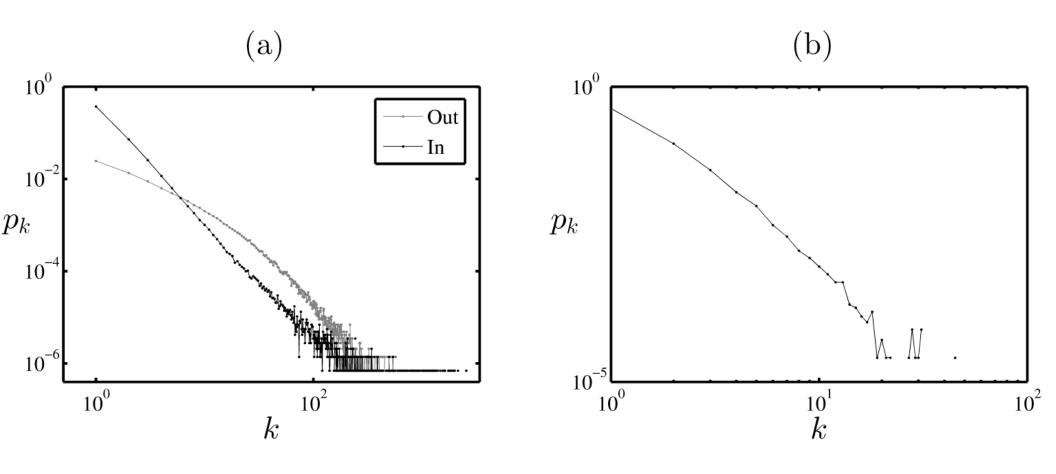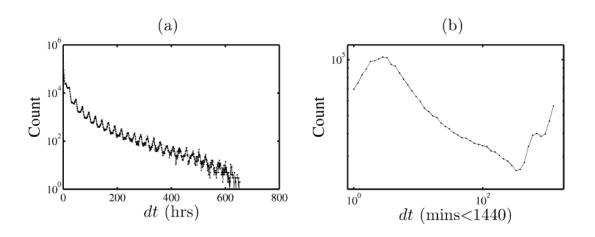
(a)

(b)

clustering coefficient

01/01   08/01

An individual daily maximum -590;
for all 28 days - 5,604;
the mean daily tweet rate -0.9410.

(a)

(b)

Conversational time-patterns

Distribution of time intervals
between consecutive tweets from
individual users

Summary

- 9hrs magic number for distinct conversations
- conversations are mostly fast-paced and balanced
- multi-user exchanges can be found thanks to data sparsity, similar threshold of 9hrs
- most exchanges dominated by 1 or 2 users
- but also some evidence of balanced (up to 5) multi-user exchange
- FFriday clustering coefficients visible

d.v.greetham@reading.ac.uk

Thanks!
d.v.greetham@reading.ac.uk

Some references:
C. Honeycutt and S. C. Herring, "Beyond microblogging: Conversation and collaboration via twitter," in Proceedings of the Forty-Second Hawai'i International Conference on System Sciences, Los Alamitos, 2009.
A. Ritter, C. Cherry, and B. Dolan, "Unsupervised modeling of twitter conversations," in HLT-NAACL 2010, 2010.
"Characterizing and curating conversation threads: expansion, focus, volume, re-entry," in Proceedings of WSDM '13. New York, NY, USA: ACM, 2013, pp. 13–22.