When Should Social Media Be Used Within The Marketing Mix?

Jon Ward, Danica Greetham, Peter Grindrod, Colin Singleton

Centre for the Mathematics of Human Behaviour + CountingLab Limited

www.countinglab.co.uk
www.reading.ac.uk/cmohb

March 13, 2013

© University of Reading 2013

How much to allocate to social media?

© University of Reading 2013

Contents

- Motivation
- Random Subgraphs
- Results
- Other Factors
- Summary

© University of Reading 2013

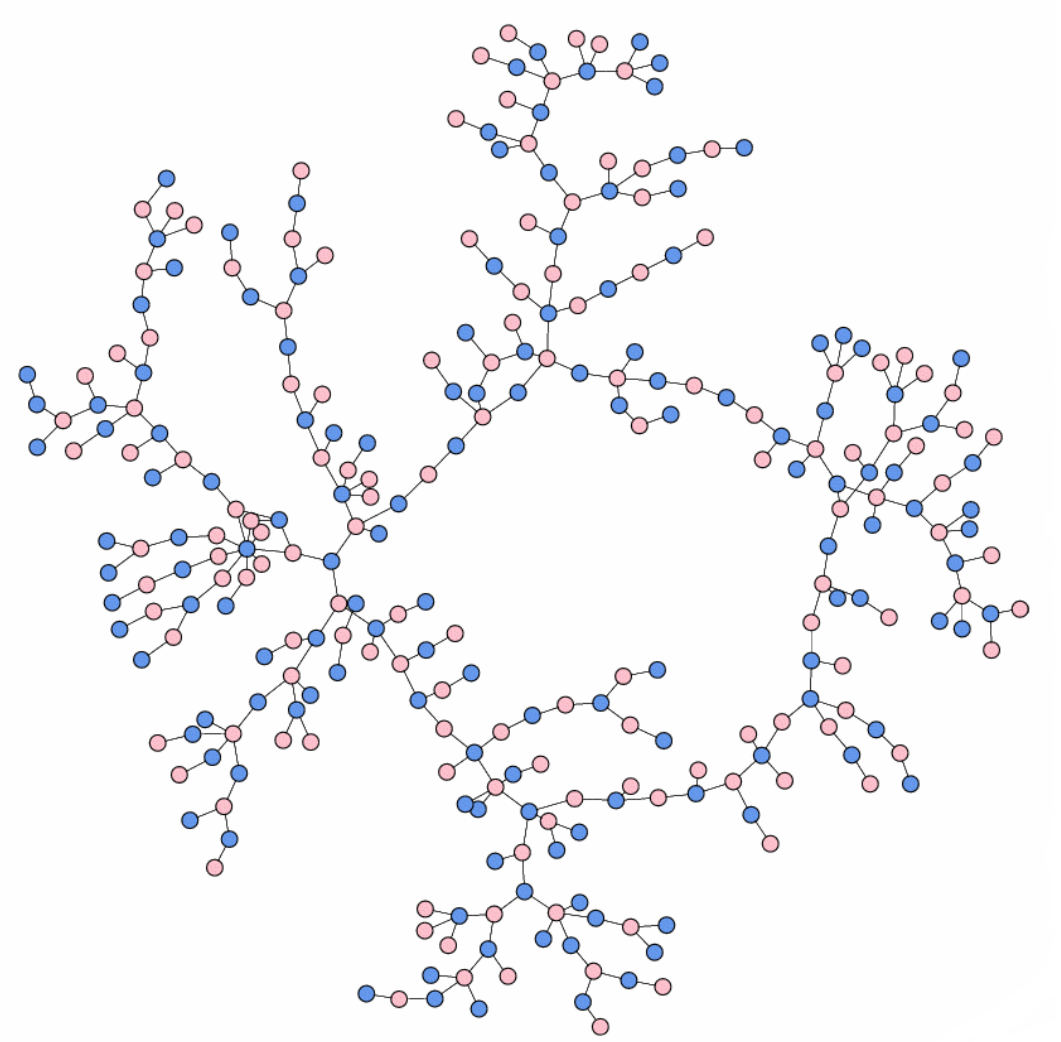
Motivation

- Companies simply do not know how much marketing spend should go towards online digital media
- Social Media marketing offers the opportunity to target individuals
- This works best when those individuals are connected and networked up – people who chat with each other through social media – knockon effects
- Some brands may have communities interested in them and therefore particularly good for social media marketing

© University of Reading 2013

Graphs

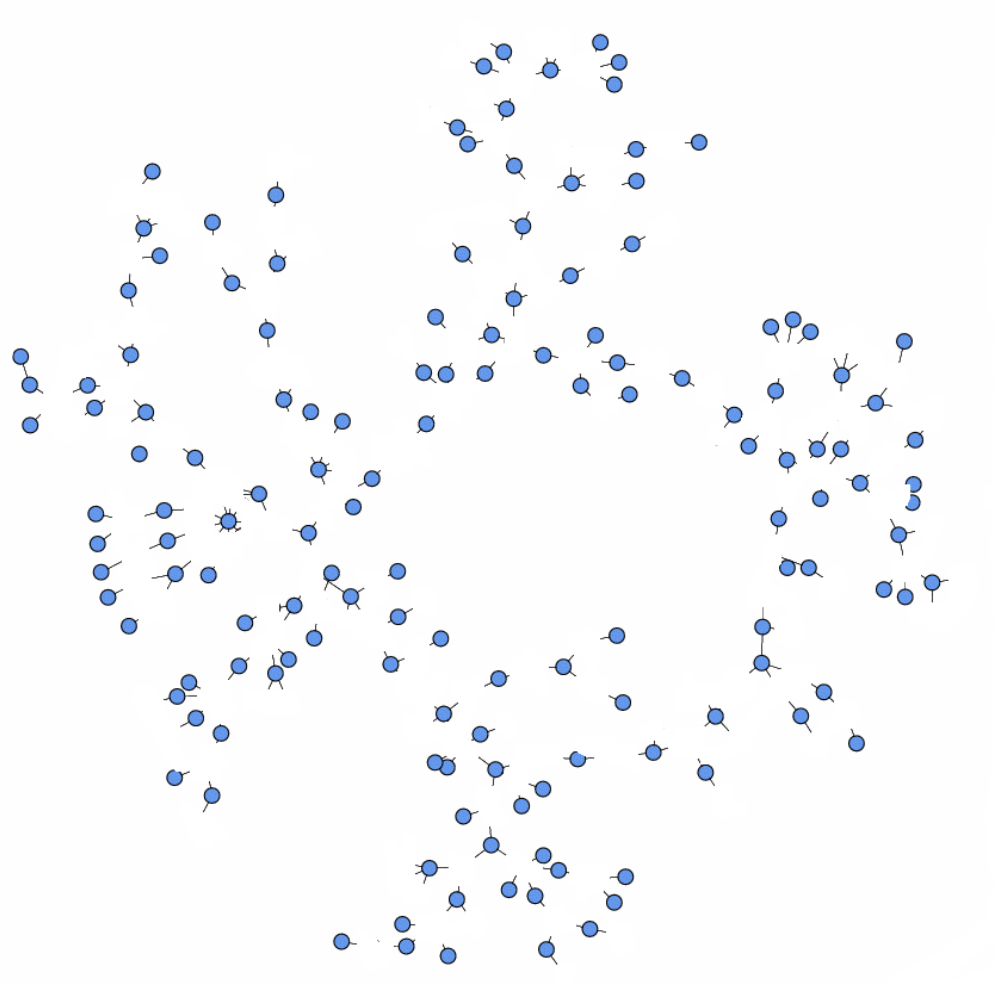
The Twitter mentions graph H(W,F) is defined by the
Nodes (Tweeters):
W = {1,….,*N*}, and
Edges (tweets with mentions)
F ð WxW



© University of Reading 2013

Subgraphs

A Subgraph G(V,E) of H is a selection V={1,…,*n*} of the original nodes W in H, along with retaining at most those edges E that were originally edges F in H and that still connect nodes in V to other nodes in V.

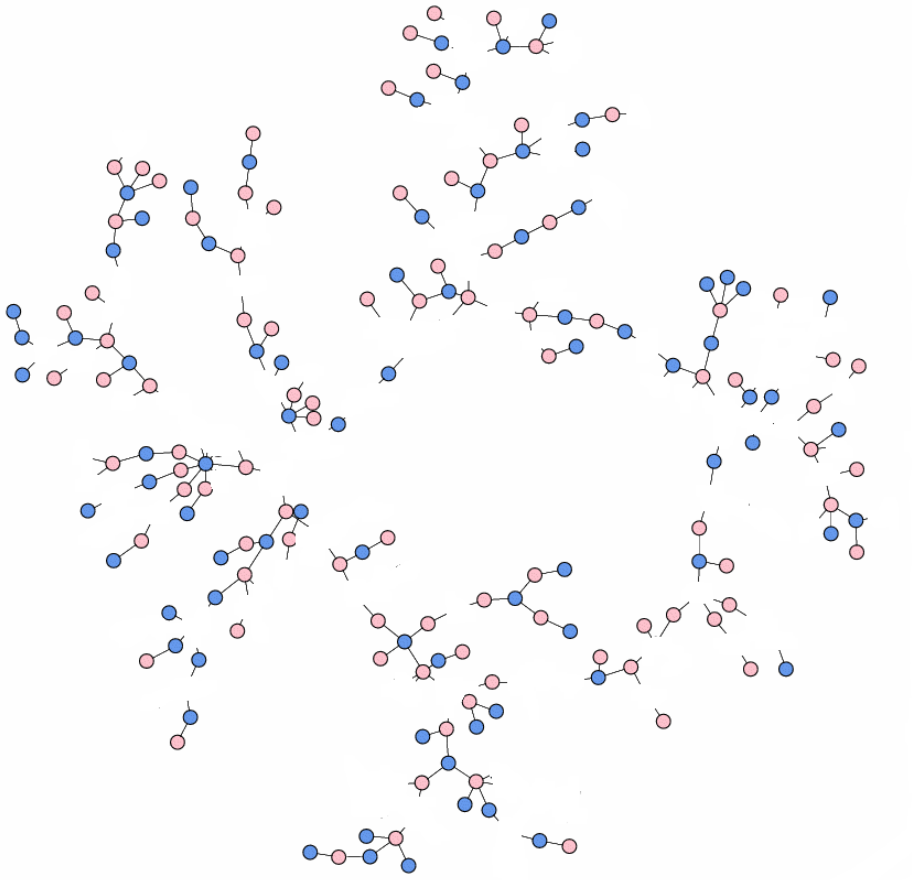


© University of Reading 2013

Random Subgraphs

Random subgraphs G(V,E) can be formed in many ways. Here we consider random subgraphs with a given number *n* nodes (V={1,…,*n*}) formed by selecting *n* nodes uniformly at random from the original graph and retaining as many edges as possible.

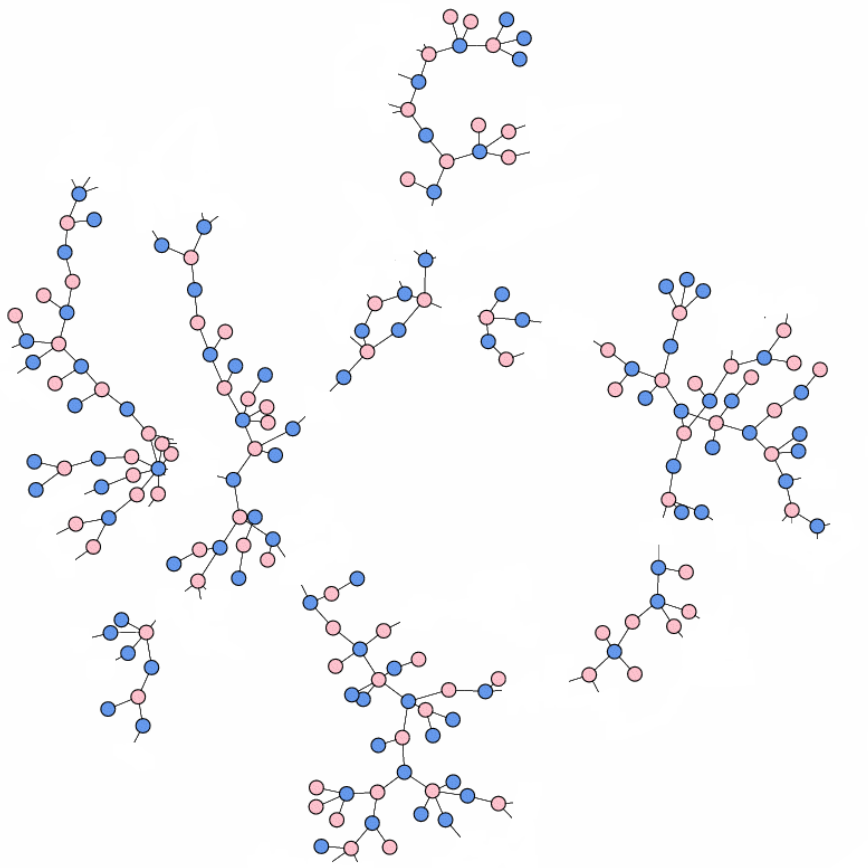
We consider the set $\mathbb{J}+\mathbb{Y}\not\mathbb{H}$, representing the set of all possible subgraphs of H that can be formed in this way.



© University of Reading 2013

Clustered Subgraphs

We might believe that a particular group of tweeters (for example, those who are fans of Partick Thistle FC) are more likely to interact through Twitter than random selection of people.
Perhaps because they are more likely to know each other, or have common interests, or because the topic of selection causes interaction online.
In this case the subgraphs may be clustered with many edges.



© University of Reading 2013

Properties of random subgraphs

- The expected value of the Watts-Strogatz clustering coefficient for the random subgraph is the same as the parent graph
- If the parent graph, H(W,F) with |W|=*N*, has degree probability distribution $\{q_0, q_1, \ldots, q_{N-1}\}$ for degrees k=0,1,…,*N*-1 then a random node on the random subgraph (of size |V| = *n*) would be expected to have degree k' with probability:

$$p_{k'} = \sum_{k=k'}^{k=k'+N-n} q_k \binom{k}{k'} \binom{N-k-1}{n-k'-1} \Big/ \binom{N-1}{n-1}$$

© University of Reading 2013

Number of edges of random subgraphs

- For large *N*, we can use that categorical distribution to produce a multinomial distribution for the degrees of each node k' = 1,…,*n* by using trials with replacement
- Note this is an approximation since one should be repeating trials *without* replacement but this can only realistically be done by numerical experiment

© University of Reading 2013

Full Graph Data

- The data comes from a tranche of public tweets from Twitter
- Sourced from Datasift
- Data is 4 weeks of public Twitter data 8th December 2011 through 4th January 2012 from users that are self-declared to be from the UK and the tweets that contain mentions (@*userB*)
- There are 4,474,693 communications from 137,084 users who sent tweets containing mentions

© University of Reading 2013

Data

- Further restriction that we create a binary undirected graph by considering that an edge exists between *i* and *j* if and only if both *i* mentions *j* and *j* mentions *i.*
- The resulting graph H(W,F) contains |W|=24,003 distinct nodes and |F|=19,452 edges.
- We also collected tweets over the same dates that contained a certain keyword (whether or not they contained mentions). Some of the keywords include: cancer, diabetes, dementia, obese, arthritis, acne, coke, stella, marmite, andrex, colgate.

© University of Reading 2013

Subgraphs Data

- From each keyword dataset, we identified the set V of all *n* users who had used that keyword in any tweet over the 4 weeks.
- We used this set V to form the keyword subgraph G(V,E).

© University of Reading 2013

Random Subgraphs

- For each size *n* of real subgraph, we generated 3.8 million simulated random subgraphs
- For each simulation a random set of *n* nodes were retained from the parent graph of *N* nodes.
- From the 3.8million random subgraphs, we counted the fraction of times the random subgraph had at least as many edges as the real subgraph giving us the "*p-v*alue".
- Where this number is very low then we can say the real subgraph contains more edges than random

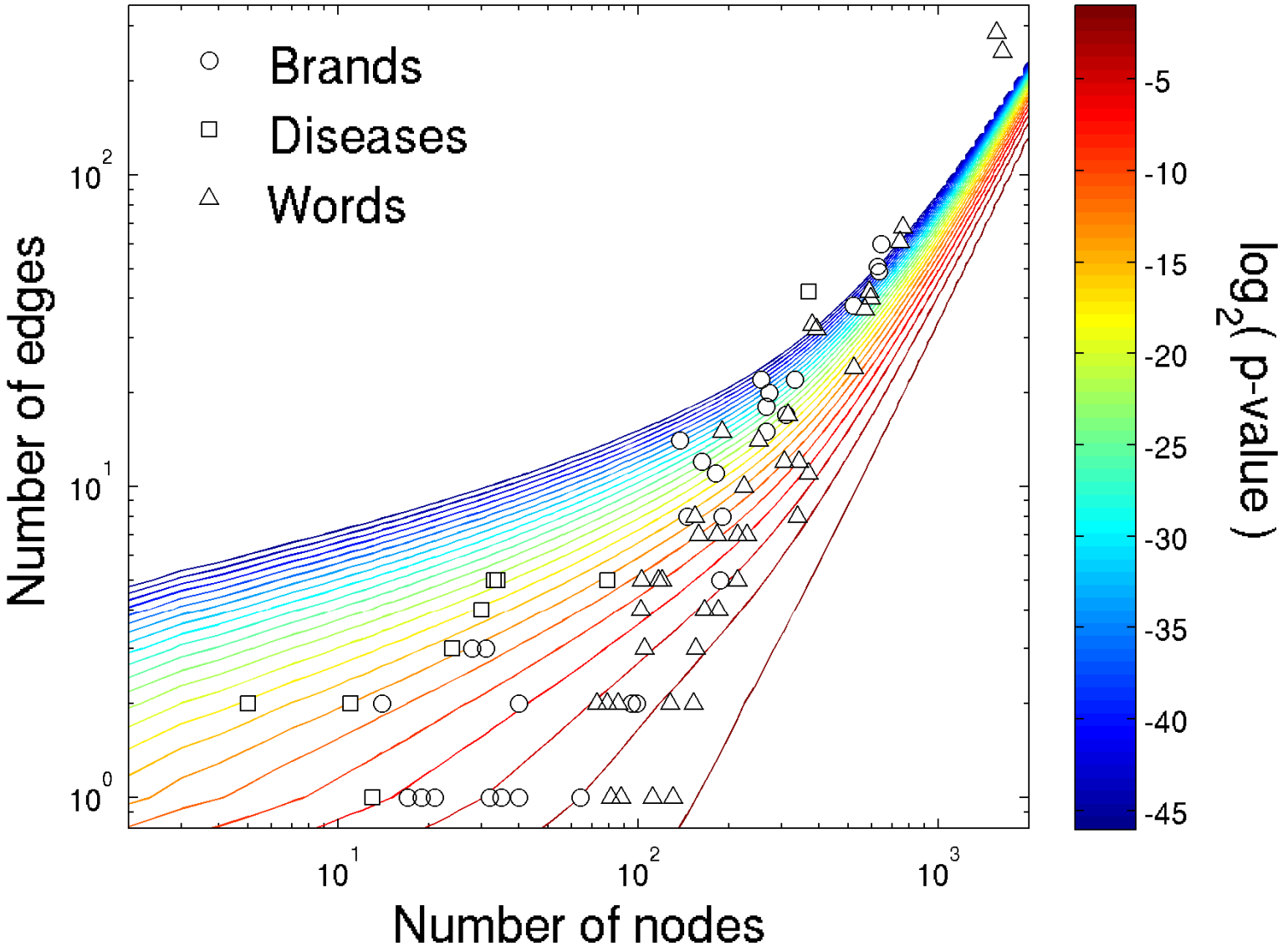
© University of Reading 2013

Random Subgraphs

- By using the categorical distribution derived earlier for the expected probability distribution of the degrees of the nodes of a subgraph with $|V|=n$ we can also determine the expected number of edges (half the expected total summed degrees).
- Further, by taking a binomial approximation to the multinomial formed by repeated trials from the categorical distribution (for these graphs typically $p_2,\ldots,p_{n-1}$ are very close to 0), or a Poisson approximation (p_1 is also typically small), one can obtain an approximate *p*-value for the number of edges without the need for numerical simulation.

© University of Reading 2013

Results



© University of Reading 2013

Health Results

Counting Lab
Centre for the Mathematics of Human Behaviour

Keyword	Number of Nodes	Number of Edges	*p*-value Numerical	Poisson	Binomial
cancer	369	42	0.00E+00	4.79E-26	6.96E-27
diabetes	33	5	0.00E+00	5.43E-10	3.97E-10
dementia	34	5	0.00E+00	7.31E-10	5.39E-10
autism	30	4	0.00E+00	3.47E-08	2.82E-08
schizophrenia	5	2	5.26E-07	3.56E-07	2.85E-07
autistic	24	3	3.42E-06	1.21E-06	1.06E-06
depression	79	5	4.47E-06	2.91E-06	2.58E-06
obese	11	2	6.05E-06	8.32E-06	7.57E-06
arthritis	13	1	5.31E-03	5.69E-03	5.69E-03
alzheimer	34	0	1.00E+00	1.00E+00	1.00E+00
acne	10	0	1.00E+00	1.00E+00	1.00E+00
asperger	8	0	1.00E+00	1.00E+00	1.00E+00
epilepsy	8	0	1.00E+00	1.00E+00	1.00E+00
dyslexia	6	0	1.00E+00	1.00E+00	1.00E+00
adhd	5	0	1.00E+00	1.00E+00	1.00E+00
asthma	5	0	1.00E+00	1.00E+00	1.00E+00
hypertension	5	0	1.00E+00	1.00E+00	1.00E+00
leukemia	4	0	1.00E+00	1.00E+00	1.00E+00
diabetic	1	0	1.00E+00	1.00E+00	1.00E+00
ocd	1	0	1.00E+00	1.00E+00	1.00E+00
anemia	0	0			
anemic	0	0			
depressant	0	0			
osteoporosis	0	0			
schizophrenic	0	0			

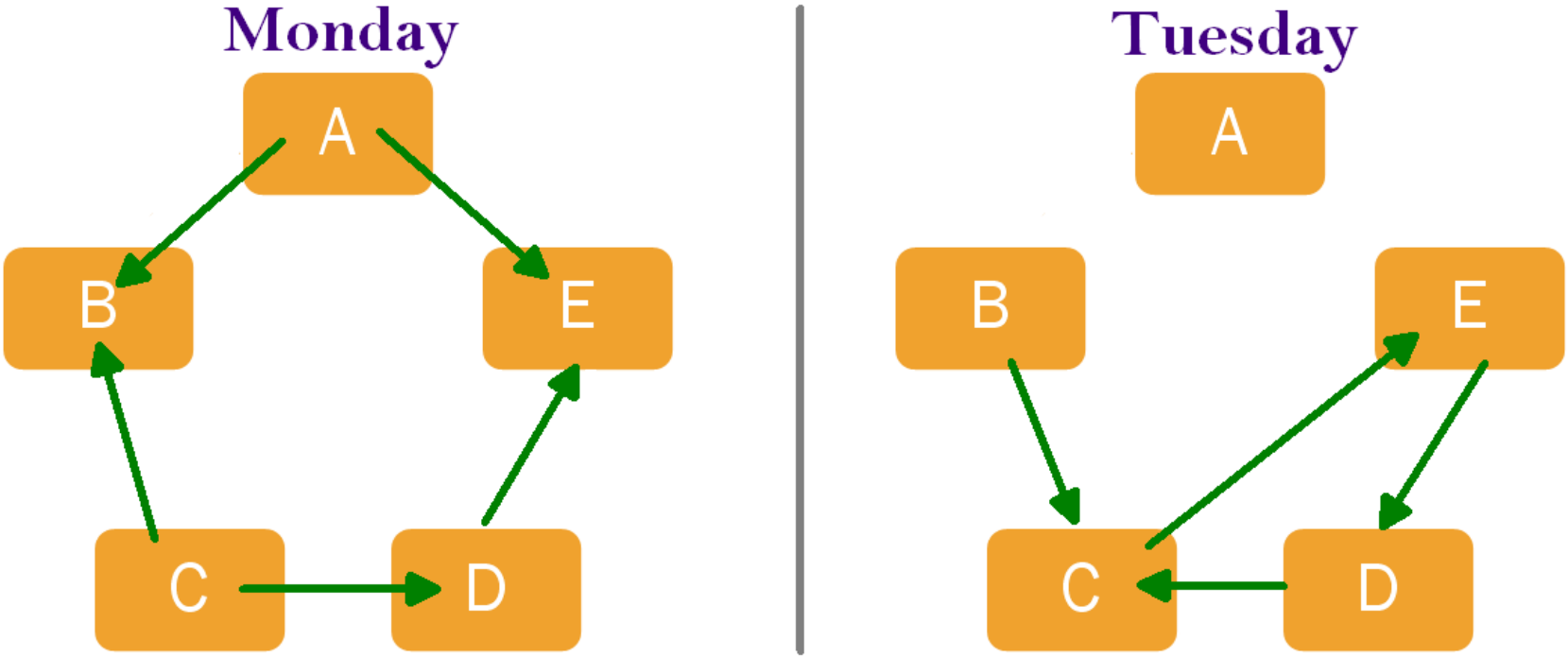
Brand Results

Keyword	Number of Nodes	Number of Edges	*p*-value Numerical	*p*-value Poisson	*p*-value Binomial
coke	2920	878	0.00E+00	1.61E-171	1.34E-202
stella	646	60	0.00E+00	7.85E-20	1.49E-20
marmite	626	51	0.00E+00	2.50E-15	8.06E-16
twix	257	22	0.00E+00	4.84E-15	2.31E-15
heineken	138	14	0.00E+00	1.30E-14	6.93E-15
pepsi	633	49	0.00E+00	8.07E-14	2.93E-14
guinness	520	38	0.00E+00	8.48E-13	3.77E-13
fosters	273	20	0.00E+00	3.90E-12	2.26E-12
malibu	333	22	0.00E+00	1.03E-10	6.30E-11
becks	267	18	0.00E+00	1.18E-10	7.60E-11
ariel	163	12	0.00E+00	2.48E-10	1.74E-10
bacardi	181	11	1.00E-06	2.77E-08	2.15E-08
carlsberg	267	15	1.00E-06	4.27E-08	3.21E-08
lynx	311	17	0.00E+00	7.14E-08	5.33E-08
corona	146	8	0.00E+00	9.44E-07	8.03E-07
twinings	28	3	6.00E-06	3.03E-06	2.72E-06
olay	31	3	8.00E-06	5.56E-06	5.04E-06
kelloggs	14	2	1.70E-05	2.18E-05	2.03E-05
jacobs	191	8	4.50E-05	4.43E-05	3.98E-05
nestle	40	2	1.35E-03	1.41E-03	1.37E-03
coca-cola	188	5	0.0079	0.0076	0.0073
febreze	17	1	0.0092	0.0097	0.0097

Keyword	Number of Nodes	Number of Edges	*p*-value Numerical	*p*-value Poisson	*p*-value Binomial
newcastle brown ale	19	1	0.012	0.012	0.012
kraft	21	1	0.014	0.015	0.015
captain morgan	32	1	0.033	0.034	0.034
gillette	95	2	0.038	0.038	0.038
frijj	35	1	0.039	0.041	0.041
tropicana	99	2	0.043	0.044	0.044
andrex	40	1	0.051	0.053	0.053
colgate	64	1	0.13	0.13	0.13
braun	74	0			
pantene	50	0			
yazoo	29	0			
charmin	27	0			
pampers	23	0			
huggies	21	0			
lipton	17	0			
persil	16	0			
domestos	15	0			
nescafe	15	0			
ambrosia	11	0			
grolsch	9	0			

© University of Reading 2013

Influence



© University of Reading 2013

Communicability

- Communicability

$$Q = \prod_{i=0}^{M} (I - \alpha A_i)^{-1}$$

I is identity matrix,

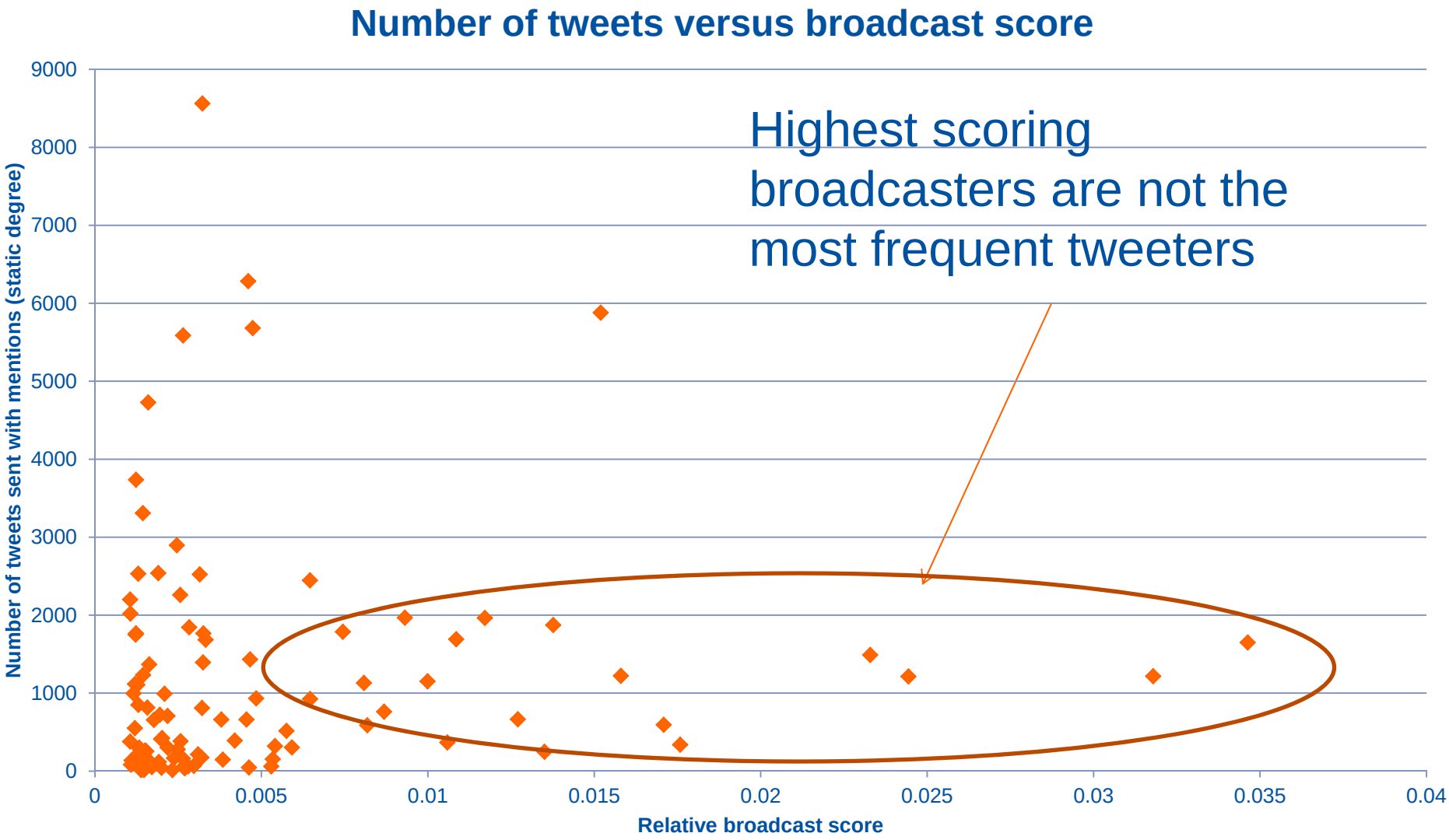
$\alpha = \frac{1}{2\max(\rho(A_i))}$, $i = 0, \cdots, M$ consecutive time-steps and

$\rho(A_i)$ is the largest eigenvalue of A_i.

Grindrod, P., Parsons, M. C., Higham, D. J. and Estrada, E. (2011) *Communicability across evolving networks.* Physical Review E, 83 (4). 046120. ISSN 1539-3755
doi: 10.1103/PhysRevE.83.046120

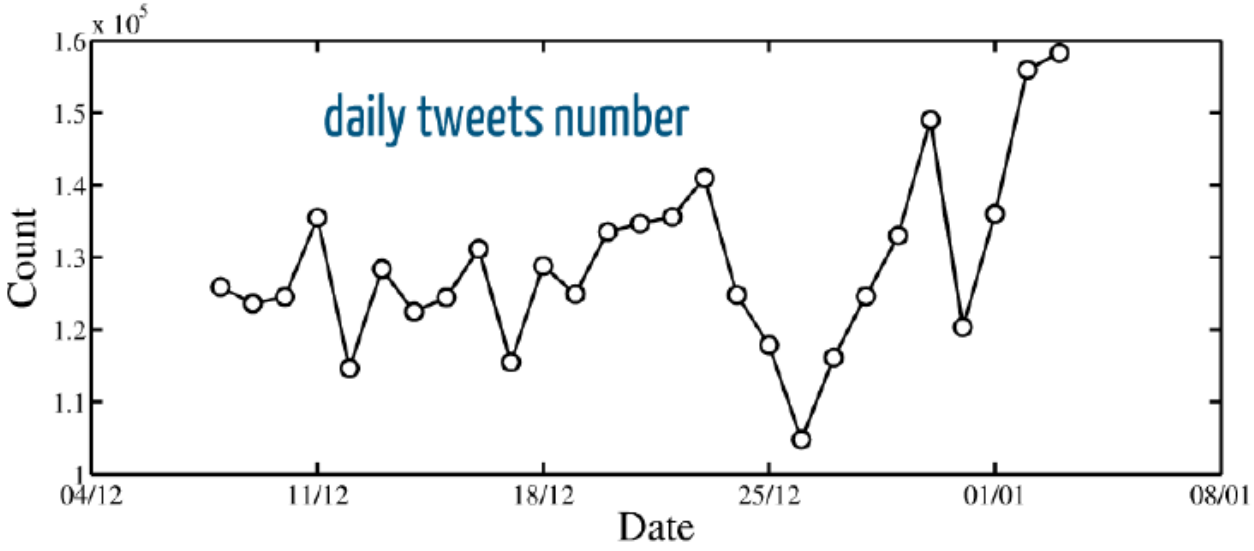
© University of Reading 2013

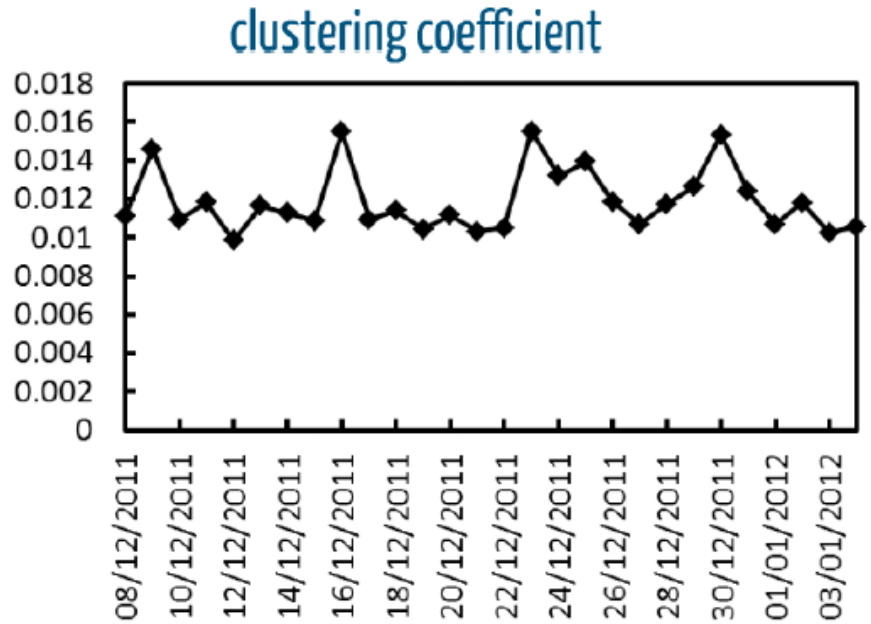
Not all about number of tweets



© University of Reading 2013

Day effects





© University of Reading 2013

Summary

- We have presented a method for producing a relative ranking of how well-connected different brands are in social media
- This provides a framework for setting a marketing budget
- Other factors to consider are how influential are the people who mention the brand
- Different days show more connected behaviour

© University of Reading 2013