TERADATA

Discovery Analytics

Chris Hillman, Principal Data Scientist (Teradata International)

Agenda

Discovery Analytics

- Level set
- Path Analytics
- Graph Analytics
- Social Sentiment

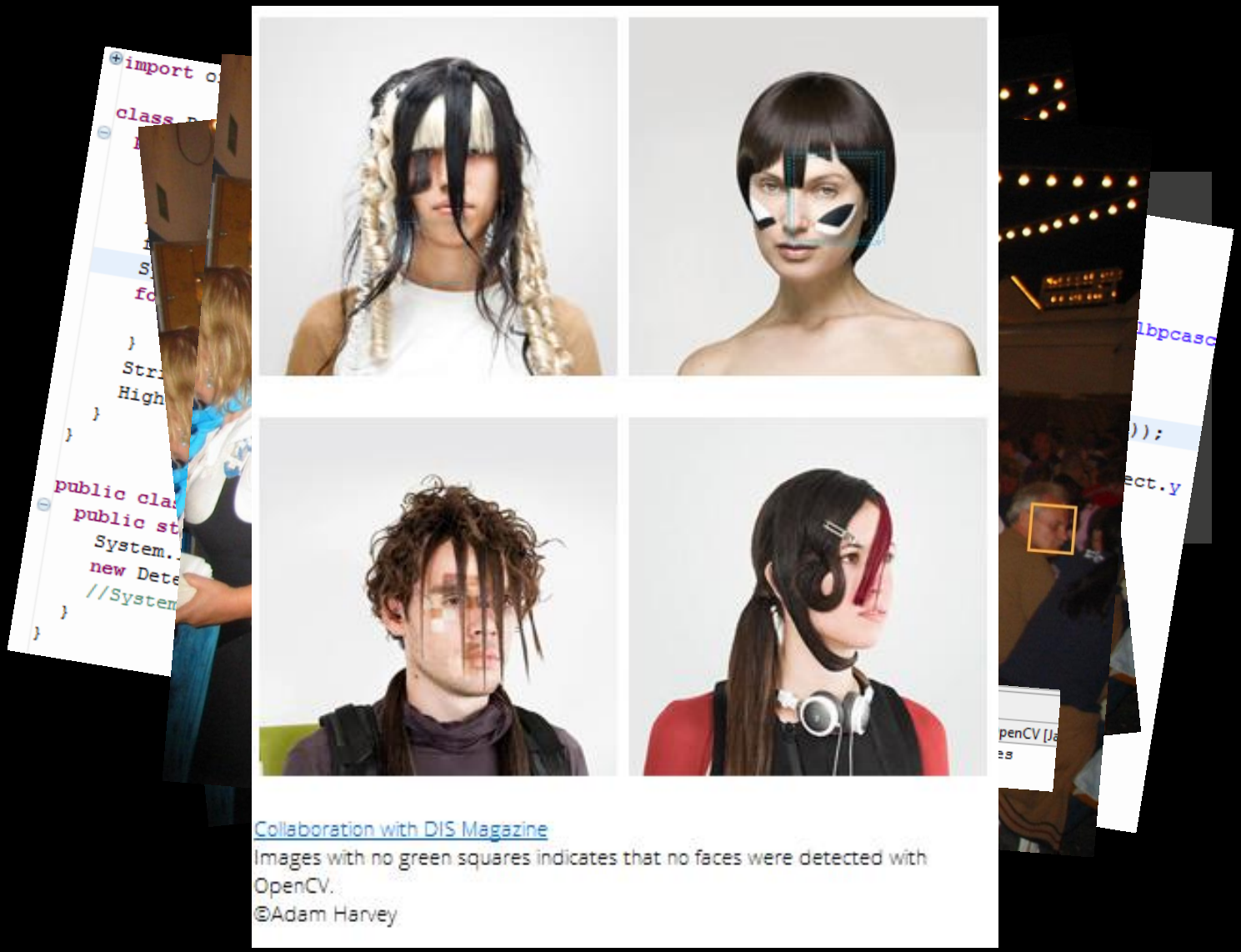
TERADATA.

What is Hadoop

http://www.flickr.com/photos/joanet/with/5797069081/

TERADATA

Face Detection – A Map Task



Collaboration with DIS Magazine
Images with no green squares indicates that no faces were detected with OpenCV.
©Adam Harvey

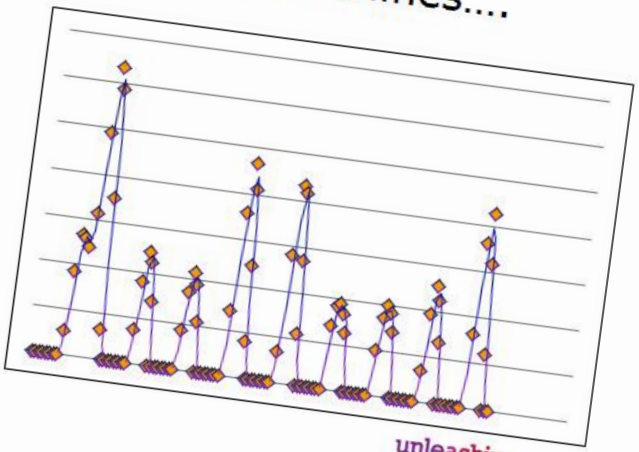
TERADATA

Proteomics

The Problem

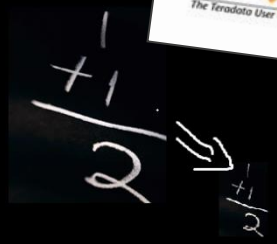
Each Experiment produces
- 5Gb XML file
- 40,000 scans
- 20,000 data points per scan
- 800,000,000 rows of data
- 2 experiments per machine
- 10 – 15 machines….

Processing the Raw data takes over 24 hours to
- Pick 2D Peaks
- De-isotope
- Pick 3D Peaks
- Match weights to known peptides



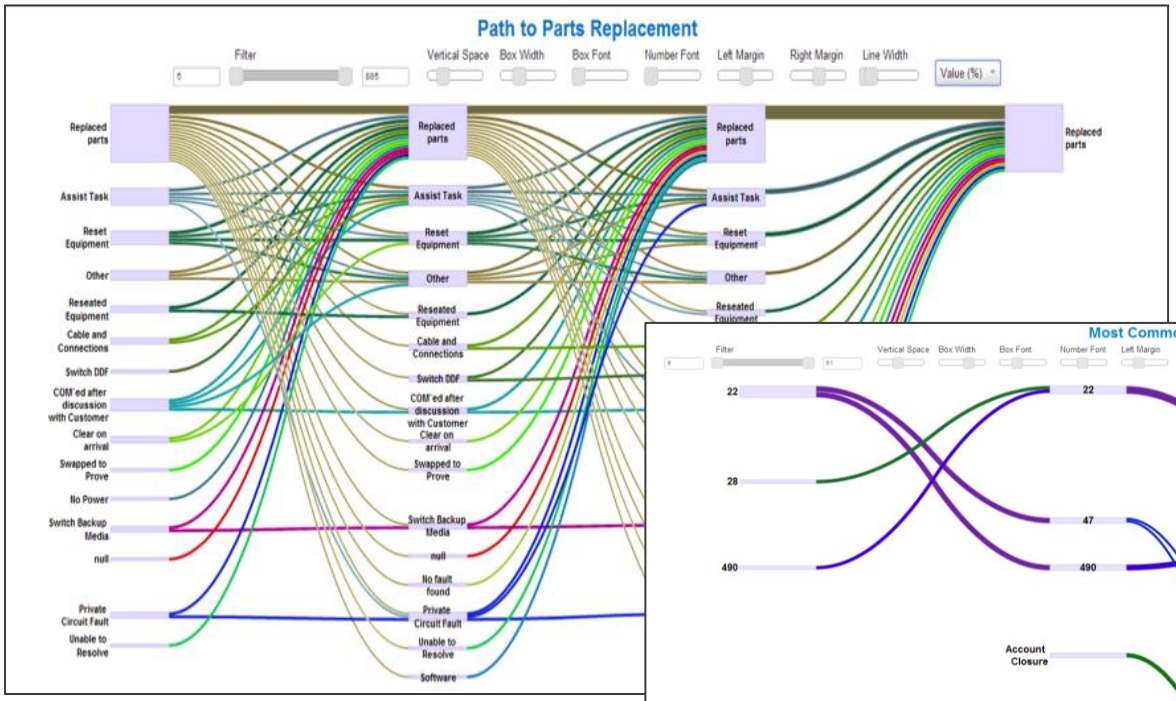
PARTNERS
The Teradata User Group

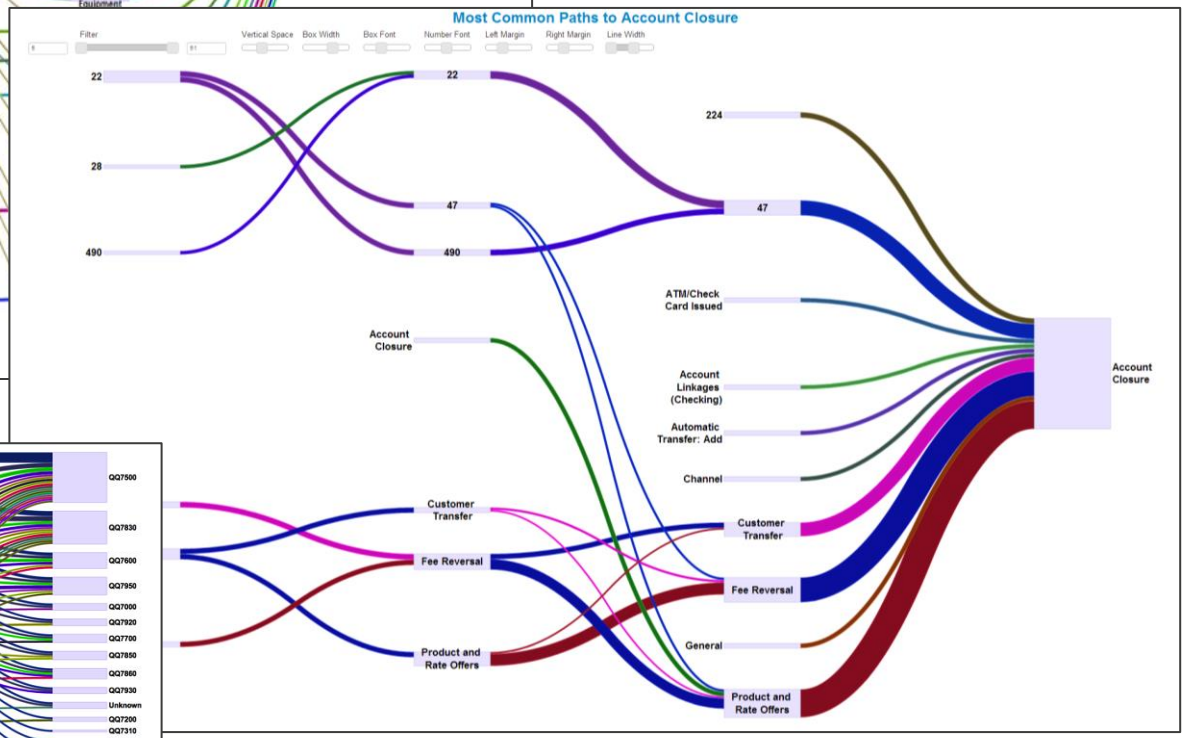
unleashing the power of DATA

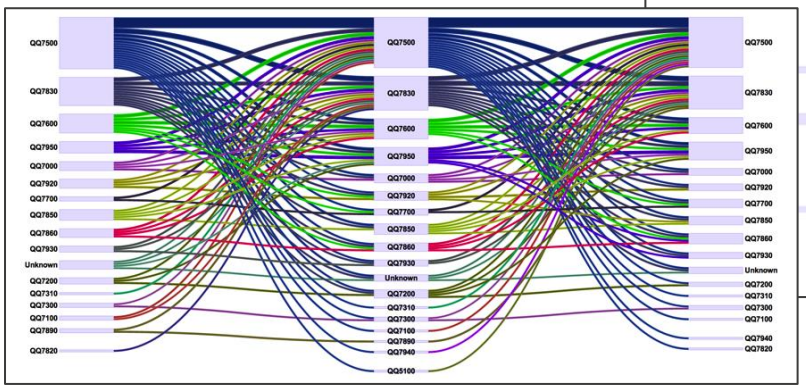


TERADATA

Patterns and paths in data

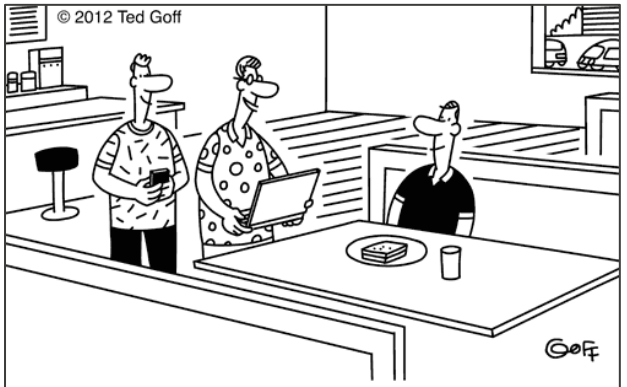




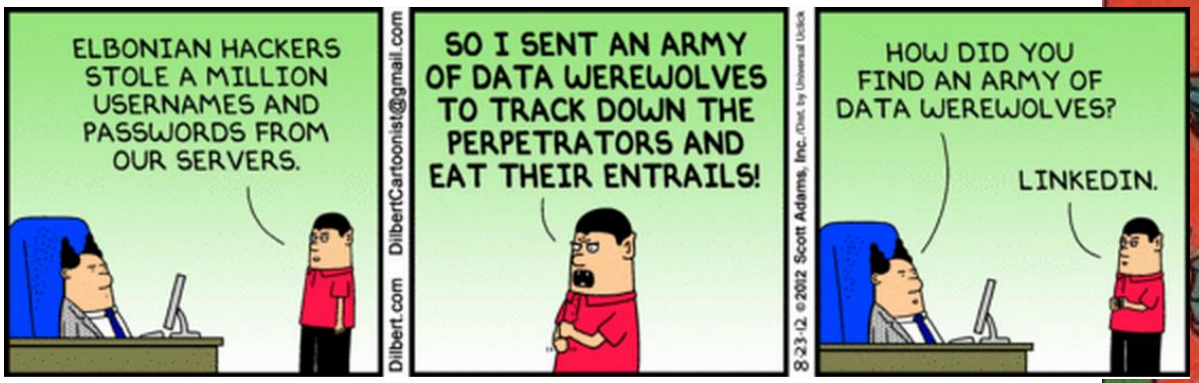


Sankey Chart used to visualise event based analytics

TERADATA

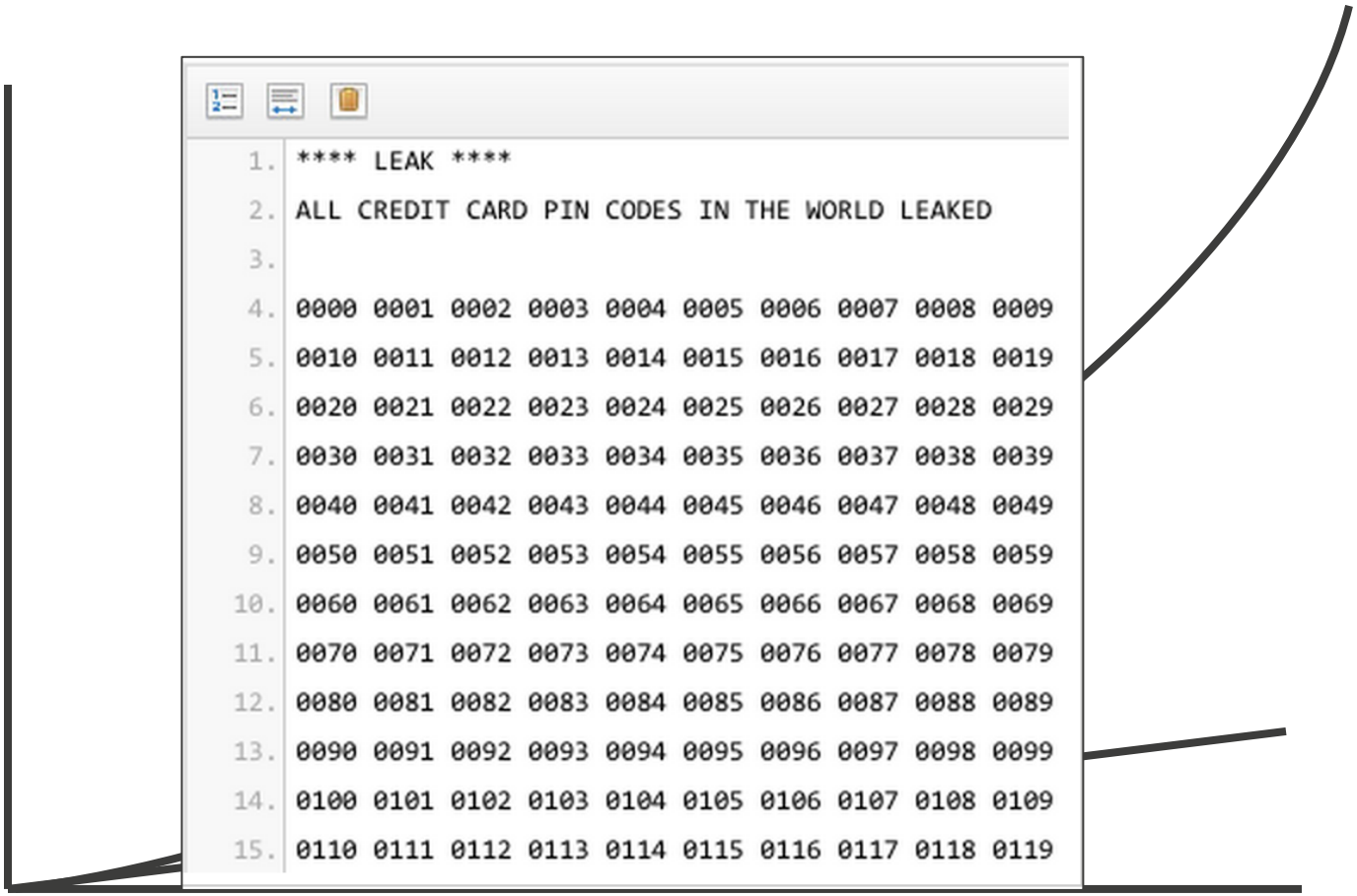


"Twitter and Facebook can't predict the election, but they did predict what you're going to have for lunch: a tuna salad sandwich. You're having the wrong sandwich."



TERADATA

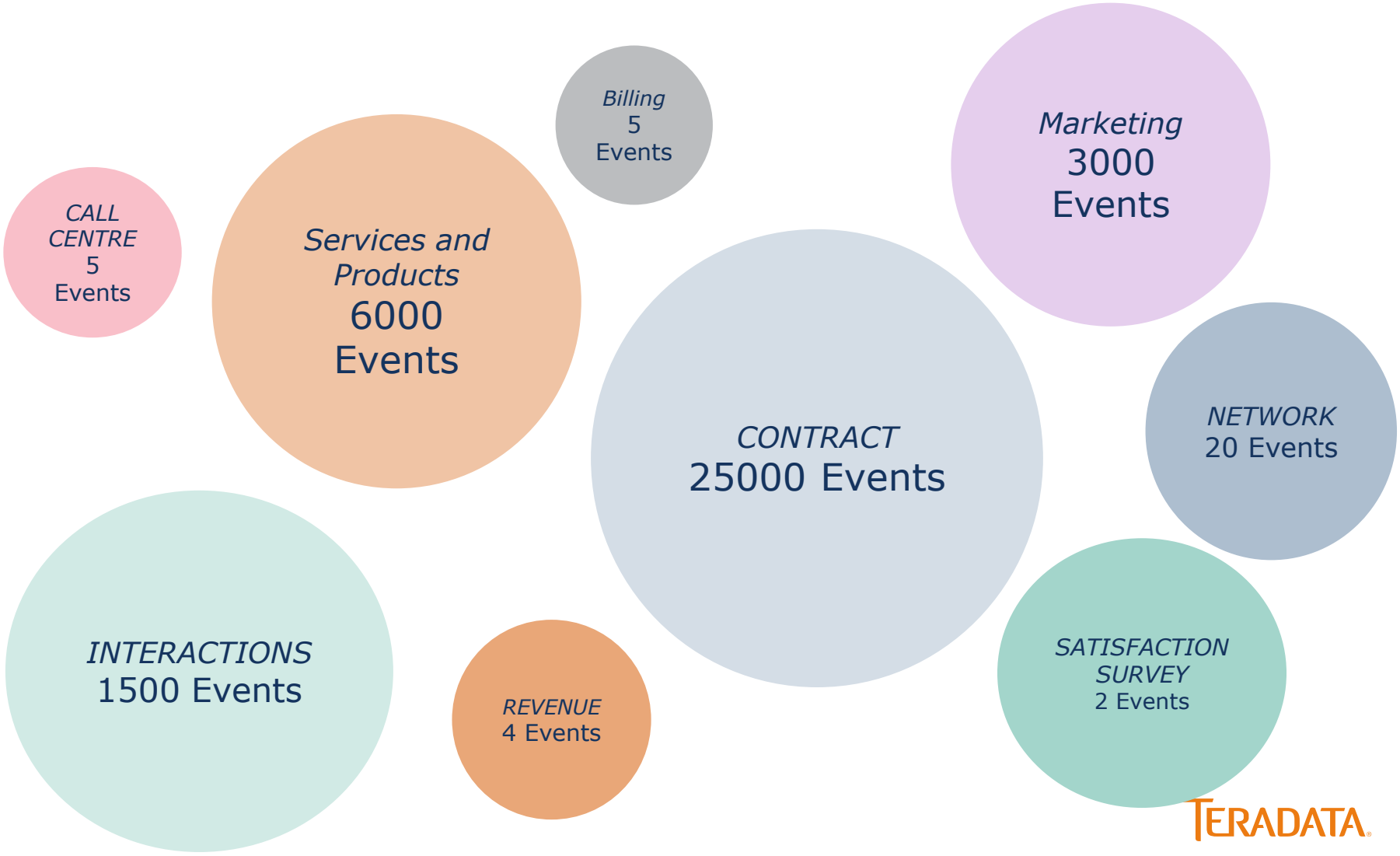
Big Data vs. Relevant Data



Event Based vs. Traditional Modeling

UID	#badcalls	ContractLength	ageband	Household	Gender	Region	married	Churned?
100681516	1	12	A	1	M	N	0	0
100681517	5	12	A	2	M	N	1	0
100681518	4	14	B	2	M	N	1	0
100681519	2	16	D	4	M	S	1	0
100681520	9	21	E	4	M	S	1	0
100681521	10	26	A	4	M	E	1	1
100681522	3	12	F	4	M	N	1	0
100681523	2	14	B	5	M	E	1	0
100681524	1	3	C	1	M	S	1	0
100681525	2	2	E	2	F	N	1	0
100681526	15	23	A	4	F	S	1	1
100681527	13	25	B	3	F	E	1	1
100681528	1	28	E	1	M	E	0	0
100681529	5	2	D	2	M	E	1	0
100681530	4	4	B	2	M	S	1	0

Event Based vs. Traditional Modeling



TERADATA

Integrated Event History Table

CONTRACT SK	DOMAIN	EVENT DETAIL	DATE & TIME
100681516	NETWORK	Bad Call Experience: Extreme	10/8/2013 13:05
100681516	NETWORK	Bad Call Experience: High	10/8/2013 13:21
100681516	PRODUCT	Subscribe to Extra Internet usage package	10/9/2013 17:50
100681516	INTERACTION	Change to online billing only	10/9/2013 17:50
100681516	CAMPAIGN	Reject offer of new Handset	10/9/2013 17:50
100681516	CAMPAIGN	Cancel Roaming package	10/12/2013 13:13
100681516	INTERACTION	Log on to online account	10/12/2013 13:14
100681516	CAMPAIGN	Reject friends and family discount	10/12/2013 14:17
100681516	INTERACTION	Complete Satisifaction survey – unhappy with dealer	10/12/2013 15:27
100681516	PRODUCT	Subscribe to free minutes offer	10/16/2013 0:00
100681516	INTERACTION	View Bill online	10/20/2013 13:19
100681516	CAMPAIGN	Reject offer of annual subscription	10/20/2013 13:20
100681516	INTERACTION	Call Centre, request contract details	10/20/2013 13:45
100681516	PRODUCT	Deactivate Extra Internet package	10/24/2013 0:00
100681516	CONTRACT	**CONTRACT - DEACTIVATED**	10/24/2013 6:45

TERADATA

Generate Triplet Patterns

CONTRACT	DOMAIN	EVENT DETAIL	DATE & TIME
100681516	NETWORK	Bad Call Experience: Extreme	10/8/2013 13:05
100681516	NETWORK	Bad Call Experience: High	10/8/2013 13:21
100681516	PRODUCT	Subscribe to Extra Internet usage package	10/9/2013 17:50
100681516	INTERACTION	Cha	
100681516	CAMPAIGN	Reje	
100681516	CAMPAIGN	Can	
100681516	INTERACTION	Log	
100681516	CAMPAIGN	Reje	
100681516	INTERACTION	Con	
100681516	PRODUCT	Sub	
100681516	INTERACTION	View	
100681516	CAMPAIGN	Reje	
100681516	INTERACTION	Call	
100681516	PRODUCT	Dea	
100681516	CONTRACT	CON	

Event 1	Event 2	Event 3	Start Date	End Date
[NETWORK] BAD CALL EXPR: EXTREME	[NETWORK] BAD CALL EXPR: MID	[NETWORK] BAD CALL EXPR: HIGH	10/8/2013 13:05	10/8/2013 13:21
[NETWORK] BAD CALL EXPR: MID	[NETWORK] BAD CALL EXPR: HIGH	[PRODUCT] INVOICE REQUEST	10/8/2013 13:13	10/9/2013 17:50
[NETWORK] BAD CALL EXPR: HIGH	[PRODUCT] INVOICE REQUEST	[INTERACTION] Change Invoice Preference	10/8/2013 13:21	10/9/2013 17:50
[INTERACTION] Change Invoice Preference	[INTERACTION] Change Invoice Preference	[CAMPAIGN] Free SMS offer	10/9/2013 17:50	10/9/2013 17:50
[INTERACTION] Change Invoice Preference	[CAMPAIGN] Free SMS offer	[PRODUCT] Roaming Minutes	10/9/2013 17:50	10/9/2013 18:07
[CAMPAIGN] Free SMS offer	[PRODUCT] Roaming Minutes	[PRODUCT] Free SMS after 6pm	10/9/2013 17:50	10/9/2013 18:07
[PRODUCT] Roaming Minutes	[PRODUCT] Free SMS after 6pm	[CAMPAIGN] New Annual Subsription	10/9/2013 18:07	10/12/2013 13:13
[INTERACTION] Change Invoice Preference	[CAMPAIGN] Free SMS offer	[INTERACTION] Change Invoice Preference	10/9/2013 18:07	10/12/2013 13:14
[CAMPAIGN] New Annual Subscription	[INTERACTION] Change Invoice Preference	[CAMPAIGN] Free SMS offer		10/12/2013 14:17

TERADATA

Predict using Naïve Bayes Classifier

EVENT 1	EVENT 2	EVENT 3	LOGLIK NON CHURN	LOGLIK CHURN	PREDICTION
[NETWORK] BAD CALL EXPR: EXTREME	[NETWORK] BAD CALL EXPR: MID	[NETWORK] BAD CALL EXPR: HIGH	-24.083	-15.5238	Churn
[NETWORK] BAD CALL EXPR: MID	[NETWORK] BAD CALL EXPR: HIGH	[PRODUCT] INVOICE REQUEST	-22.4175	-13.9183	Churn
[NETWORK] BAD CALL EXPR: HIGH	[PRODUCT] INVOICE REQUEST	[INTERACTION] Change Invoice Preference	-25.1901	-16.6981	Churn
[INTERACTION] Change Invoice Preference	[INTERACTION] Change Invoice Preference	[CAMPAIGN] Free SMS offer	-25.8888	-15.5558	Churn
[NETWORK] BAD CALL EXPR: MID	[NETWORK] BAD CALL EXPR: HIGH	[CONTRACT] 2 Year Subscription	-18.4708	-18.4708	Churn
[NETWORK] BAD CALL EXPR: HIGH	[PRODUCT] INVOICE REQUEST	[CONTRACT] 2 Year Subscription	-18.4708	-18.4708	Churn
[CAMPAIGN] Accept Upgrade	[PRODUCT] Transfer balance to Prepay	[NETWORK] BAD CALL EXPR: HIGH	-23.1044	-29.2531	Non Churn
[NETWORK] BAD CALL EXPR: MID	[CAMPAIGN] Accept Upgrade	[CONTRACT] 2 Year Subscription	-24.4797	-29.2759	Non Churn
[CONTRACT] 2 Year Subscription	[CAMPAIGN] Accept Upgrade	[INTERACTION] Change Invoice Preference	-22.7526	-27.3709	Non Churn
[CONTRACT] 2 Year Subscription	[CAMPAIGN] Accept Upgrade	[CAMPAIGN] Free SMS offer	-17.4364	-22.0436	Non Churn
[INTERACTION] Change Invoice Preference	[INTERACTION] Change Invoice Preference	[CAMPAIGN] Free SMS offer	-19.7834	-24.2644	Non Churn

Naïve Bayes makes a prediction as to whether a triplet pattern is a churn/non-churn triplet. These triplets can then be compared to contract journeys to make a determination of whether a contract is likely to churn.

TERADATA

Churn Contract Predictions

Contract ID	Churn Predictions	NonChurn Predictions	Churn Time Scores	NonChurn Time Scores	Churn Freq. Score	Churn Time Score	Prediction	Actual
1	23	8	227	80	2.875	2.8375	Churn	Churn
2	12	7	119	63	1.71429	1.88889	Churn	Churn
3	8	2	65	20	4	3.25	Churn	Churn
4	23	11	218	105	2.09091	2.07619	Churn	Churn
5	10	2	92	19	5	4.84211	Churn	Churn
6	24	19	230	190	1.26316	1.21053	Churn	Churn
7	14	3	129	30	4.66667	4.3	Churn	Churn
8	16	8	141	80	2	1.7625	Churn	Churn
9	6	2	57	20	3	2.85	Churn	Churn

The time duration of a triplet is used to adjust scores.

If triplets occurred over a longer period of time it has a negative affect on the score, versus if the triplet occurred over a short period of time.

Contract ID	Churn Predictions	NonChurn Predictions	Churn Time Scores	NonChurn Time Scores	Churn Freq. Score	Churn Time Score	Prediction	Actual
10	2	11	14	77	0.181818	0.153846	Not Churn	Not Churn
11	3	18	30	152	0.166667	0.197368	Not Churn	Not Churn
12	9	92	90	914	0.097826	0.098468	Not Churn	Not Churn
13	2	15	20	134	0.133333	0.149254	Not Churn	Not Churn
14	2	10	20	79	0.2	0.253165	Not Churn	Not Churn
15	4	7	40	63	0.571429	0.634921	Not Churn	Not Churn
16	5	8	47	71	0.625	0.661972	Not Churn	Not Churn
17	5	38	45	356	0.131579	0.126404	Not Churn	Not Churn

TERADATA

Measure Effectiveness

MODEL SENSITIVITY

Measure of how accurate the model is at predicting churners (True Positives).

$$\text{Sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

70%

MODEL SPECIFICITY

Measure of how accurate the model is at predicting non-churners (True Negatives).

$$\text{Specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}$$

95%

MODEL LIFT

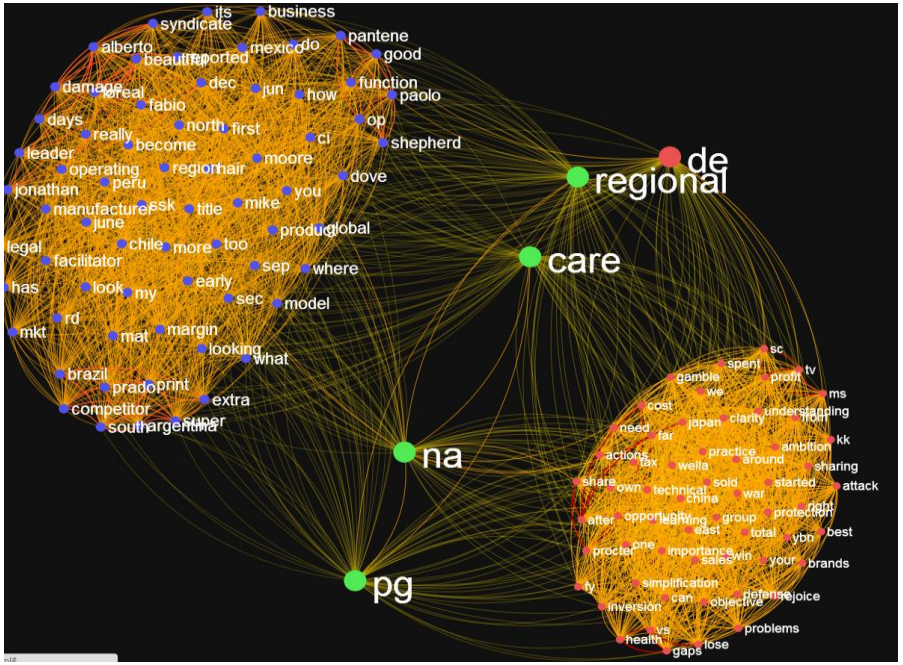
Indicator of model quality and performance.

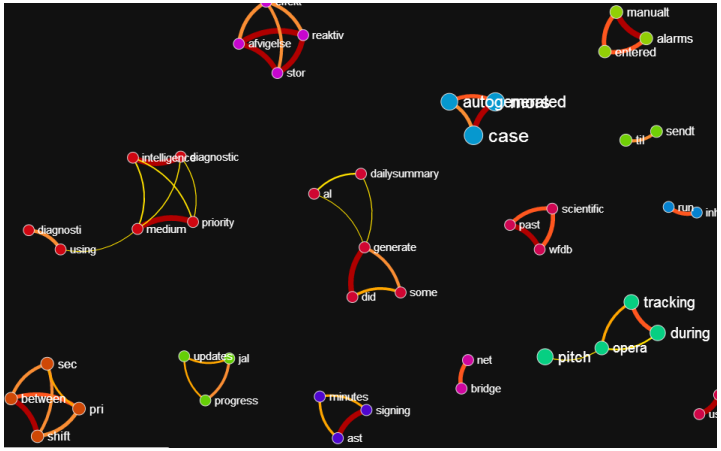
Churners detected in Top Decile by Aster: 100k

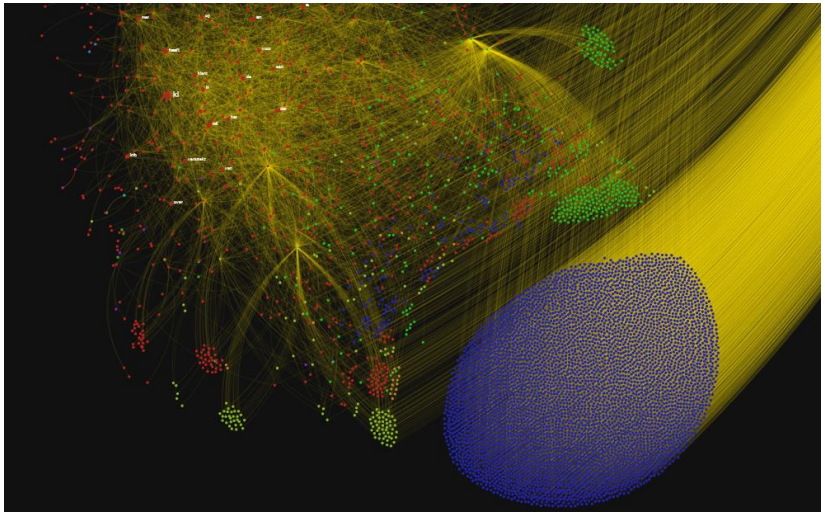
6

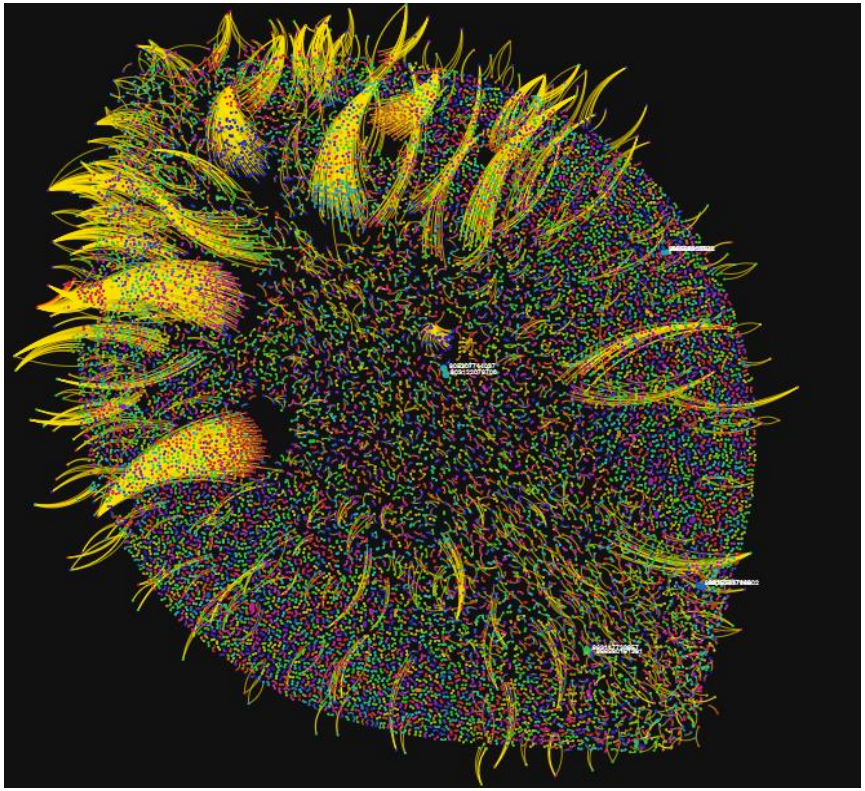
TERADATA

Graphs





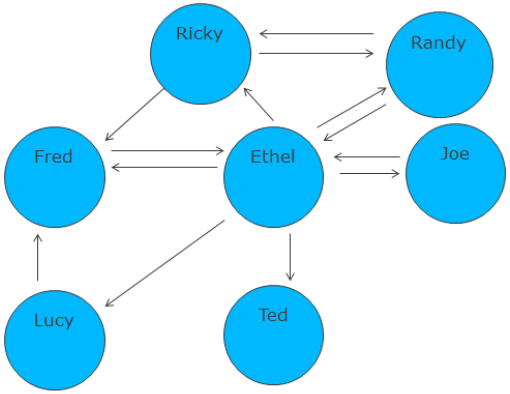




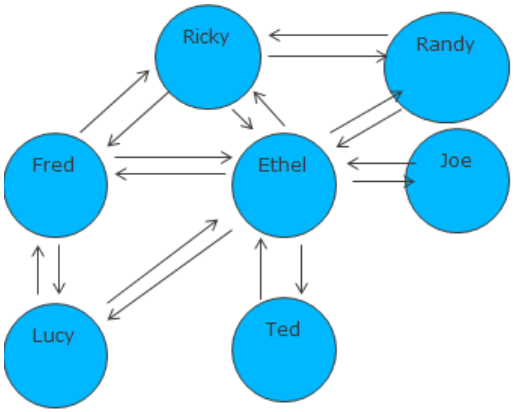
TERADATA

Graph Metrics

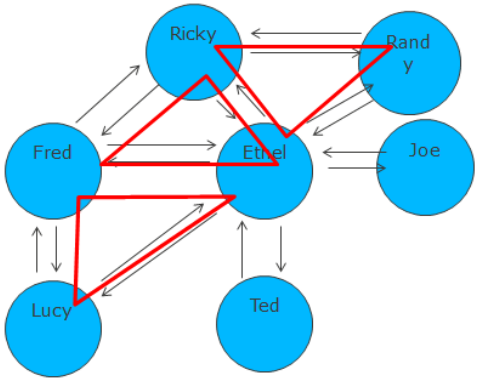
Directed Graph



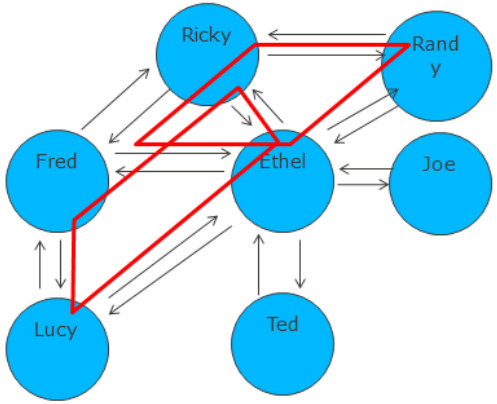
Degree Centrality



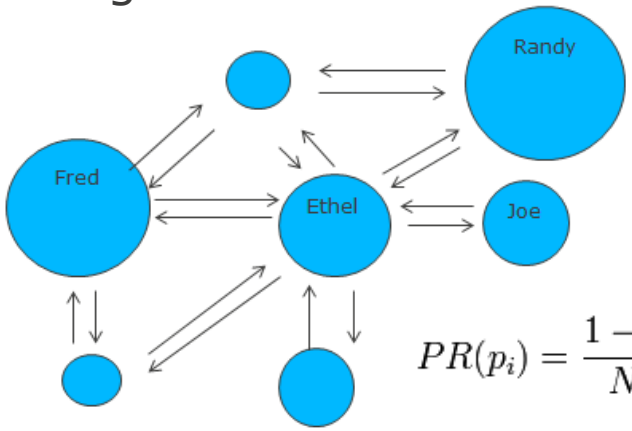
Triangle Counts



Rectangle Counts



Pagerank



$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

TERADATA.

Building a graph metric table

- Using Simple SQL joins to put the pieces together into a table by user by month with the graph metrics as columns. The Graph metrics such as pagerank, centrality and triangle counts add to the understanding and importance of interactions between users

Creators	Views	Outdegree	Indegree	Pagerank	Triangle Count	Rectangle Count
jetsonpete	406	309	14	0.007460	116	440
Vazquex9999	1043	363	11	0.004323	377	1044
Squire45	116	235	10	0.0042112	22	543
Chillax7	643	135	10	0.0042111	76	235
mumford	43	96	10	0.0041987	45	132
sh4rker	183	32	65	0.0041878	12	122
ibanexxx	186	88	12	0.0041877	111	299
baddevotions	908	62	11	0.0041442	253	743
harris1	130	63	9	0.0040345	332	221
Bobster	123	63	14	0.007460	211	223
Harebrain12	122	130	10	0.0041878	32	432
Ibanexxxx	403	211	12	0.0041877	89	121
T3l3c4st3r	221	92	11	0.0041442	32	221
DrRock	225	12	15	0.0040345	118	543

Compare Users by simple and Graph metrics

Creators	July 13	Views	July 13	Viewers	July 13	Centrality	July13	PageRank	July13
plumbago	85	Daveyboy	1483	Sh4rker	500	Smithy	0.2833	Sh4rker	0.007460
Vazquex9999	69	Ibanexxx	812	Lospollos	158	Tunf	0.2813	Nicky	0.004323
Squire45	49	Liza8	664	Frankenfurt	151	Harris1	0.2515	Tommo14	0.0042112
Chillax7	49	Frankenfurt	557	Ibanexxx	147	Ibanexxx	0.2247	Mrjeffreys	0.0042111
Daveyboy	49	Gill	515	jetsonpete	132	Jetsonpete	0.2238	Jetsonpete	0.0041987
jetsonpete	47	Smithy	380	Daveyboy	128	Bobster	0.2076	Dougcutting	0.0041878
Ibanexxx	42	Baddevotions	378	Jessie	127	Martinez	0.1903	Ffdsa	0.0041877
Baddevotions	38	Jetsonpete	330	Bluerock	121	Massy69	0.1785	Lydia1	0.0041442
Sh4rker	32	63danny	288	Floyd911	1			d911	0.0040345
Bobster	31	Martinez	282	Evilgus	1			32	0.007460
Tweedy	30	Johnny	270	Heisenburg	1			mandeveit	0.0042112
Canil	29	Jeff.master	214	Baddevotions	1			re45	0.0042111
Dotchin	27	maclad	210	Harebrain12	1			gus	0.0041987
Harebrain12	26	Harris1	200	Mustang899	9			tang899	0.0041878
Ibanexxxx	24	Bobster	192	Chillax7	98	Mrjeffreys	0.1337	Phillipjones	0.0041877
T3l3c4st3r	23	Ianburton	172	Lydia1	92	Tscore	0.1335	Harebrain12	0.0041442
DrRock	21	Chaveyjax	163	Bobster	92	Iceman	0.1319	Normanbates	0.0040345
Chaveyjax	20	Tivertom	160	90082	90	Kingkong	0.1291	Yandt	0.0032122
Tunf	19	Tunf	159	Tivertom	71	Zola21	0.1288	Windbag	0.0031122
Mailman	18	Weareborg	156	Clarky	69	Nikkisixx	0.1265	Torres	0.0021265
mumford	18	chilla	129	venus12	68	Mckeith	0.1213	Whatssup	0.0021092

Scores Highly on simple count metrics but does not appear in top Graph Metrics

Do you like Football?

Timestamp	User	Activity	Type	Subject
4/2/2013 11:03:51	Daveyboy	Create	Thread	En-ger-land
4/16/2013 8:52:20	Daveyboy	Create	Thread	World cup fantasy football
4/1/2013 1:52:10	Daveyboy	Create	Thread	Lampard or Gerrard?
4/22/2013 2:03:19	Daveyboy	Create	Message	World cup fantasy footb
4/18/2013 10:25:00	Daveyboy	Create	Message	World cup fantasy footb
4/9/2013 11:03:51	Daveyboy	Create	Thread	Thoughts on Wednesda
4/17/2013 11:01:35	Daveyboy	Create	Message	World cup fantasy footb
4/22/2013 1:47:49	Daveyboy	Create	Message	Lampard or Gerrard?
4/2/2013 9:13:42	Daveyboy	Create	Video	Lampard or Gerrard?
4/24/2013 12:45:27	Daveyboy	Create	Message	Thoughts on Wednesday's game
4/22/2013 4:00:40	Daveyboy	Create	Message	Thoughts on Wednesday's game
4/16/2013 4:43:05	Daveyboy	Create	Message	Thoughts on Wednesday's game
4/9/2013 12:40:02	Daveyboy	Create	Message	Lampard or Gerrard?
4/3/2013 8:52:20	Daveyboy	Create	Message	Thoughts on Wednesday's game
4/16/2013 6:40:38	Daveyboy	Create	Video	Thoughts on Wednesday's game
4/24/2013 9:20:11	Daveyboy	Create	Message	World cup fantasy football
4/25/2013 8:18:49	Daveyboy	Create	Video	World cup fantasy football
4/23/2013 3:15:05	Daveyboy	Create	Message	World cup fantasy football
4/1/2013 10:24:40	Daveyboy	Create	Thread	playoffs
4/9/2013 5:21:26	Daveyboy	Create	Message	Thoughts on Wednesday's game
4/1/2013 2:11:11	Daveyboy	Create	Message	playoffs

Lots of posts and views but all on a single topic

TERADATA

Compare Users by simple and Graph metrics

Creators	July 13	Views	July 13	Viewers	July 13	Centrality	July13	PageRank	July13
plumbago	85	Daveyboy	1483	Sh4rker	500	Smithy	0.2833	Sh4rker	0.007460
Vazquex9999	69	Ibanexxx	812	Lospollos	158	Tunf	0.2813	Nicky	0.004323
Squire45	49	Liza8	664	Frankenfurt	151	Harris1	0.2515	Tommo14	0.0042112
Chillax7	49	Frankenfurt	557	Ibanexxx	147	Ibanexxx	0.2247	Mrjeffreys	0.0042111
Daveyboy	49	Gill	515	jetsonpete	132	Jetsonpete	0.2238	Jetsonpete	0.0041987
jetsonpete	47	Smithy	380	Daveyboy	128	Bobster	0.2076	Dougcutting	0.0041878
Ibanexxx	42	Baddevotions	378	Jessie	127	Martinez	0.1903	Ffdsa	0.0041877
Baddevotions	38	Jetsonpete	330	Bluerock	121	Massy69	0.1785	Lydia1	0.0041442
Sh4rker	32	63danny	288	Floyd911	118	Meerkat	0.1447	Floyd911	0.0040345
Bobster	31	Martinez	282	Evilgus	115	Langster	0.1440	90082	0.007460
Harris1	30	Johnny	270				0.1438	Adamandeveit	0.0042112
Canil	29	Jeff.master	214				0.1403	Squire45	0.0042111
Dotchin	27	maclad	210	Harebrain12	101	Neilneil	0.1366	Evilgus	0.0041987
Harebrain12	26	Harris1	200	Mustang899	99	Deacona	0.1344	Mustang899	0.0041878
Ibanexxxx	24	Bobster	192	Chillax7	98	Mrjeffreys	0.1337	Phillipjones	0.0041877
T3l3c4st3r	23	Ianburton	172	Lydia1	92	Tscore	0.1335	Harebrain12	0.0041442
DrRock	21	Chaveyjax	163	Bobster	92	Iceman	0.1319	Normanbates	0.0040345
Chaveyjax	20	Tivertom	160	90082	90	Kingkong	0.1291	Yandt	0.0032122
Tunf	19	Tunf	159	Tivertom	71	Zola21	0.1288	Windbag	0.0031122
Mailman	18	Weareborg	156	Clarky	69	Nikkisixx	0.1265	Torres	0.0021265
mumford	18	chilla	129	venus12	68	Mckeith	0.1213	Whatssup	0.0021092

High Centrality but low Page Rank

TERADATA

Not interested in other peoples posts?

User	Activity	Subject
Harris1	Create	Swimmers call to action
Harris1	Create	Anyone using excel macros
Harris1	Create	Ironman trophy
Harris1	Create	Program management kickoff
Harris1	Create	Unacceptable meeting bookings
Harris1	Create	Off topic but I agree
Harris1	Create	Tickets for club night on sale tomorrow
Harris1	Create	Kit list for lunchtime workouts
Harris1	Create	Travelcard or Oyster card
Harris1	Create	What's the best way to contact IT support no
Harris1	Create	My thoughts on the organisation changes
Harris1	Create	House prices rocketing in some areas
Harris1	Create	We need to all stop this now!
Harris1	Create	Smoking outside open windows
Harris1	Create	Project management best practices attached
Harris1	Create	Please shut the doors of meeting rooms
Harris1	Create	Measuring success
Harris1	View	My thoughts on the organisation changes
Harris1	View	What's the best way to contact IT support nowadays
Harris1	View	Unacceptable meeting bookings

High Number of Posts on a variety of topics which are have a low number of views… also low number of views of other peoples posts

TERADATA

Compare Users by simple and Graph metrics

Creators	July 13	Views	July 13	Viewers	July 13	Centrality	July13	PageRank	July13
plumbago	85	Daveyboy	1483	Sh4rker	500	Smithy	0.2833	Sh4rker	0.007460
Vazquex9999	69	Ibanexxx	812	Lospollos	158	Tunf	0.2813	Nicky	0.004323
Squire45	49	Liza8	664	Frankenfurt	151	Harris1	0.2515	Tommo14	0.0042112
Chillax7	49	Frankenfurt	557	Ibanexxx	147	Ibanexxx	0.2247	Mrjeffreys	0.0042111
Daveyboy	49	Gill	515	jetsonpete	132	Jetsonpete	0.2238	Jetsonpete	0.0041987
jetsonpete	47	Smithy	380	Daveyboy	128	Bobster	0.2076	Dougcutting	0.0041878
Ibanexxx	42	Baddevotions	378	Jessie	127	Martinez	0.1903	Ffdsa	0.0041877
Baddevotions	38	Jetsonpete	330	Bluerock	121	Massy69	0.1785	Lydia1	0.0041442
mumford	32	63danny	288	Floyd91				Floyd911	0.0040345
Bobster	31	Martinez	282	Evilgus				90082	0.007460
Harris1	30	Johnny	270	Heisenburg	104	Alicewhitey	0.1438	Adamandeveit	0.0042112
Canil	29	Jeff.master	214	Baddevotions	103	Sarahc	0.1403	Squire45	0.0042111
Dotchin	27	maclad	210	Harebrain12	101	Neilneil	0.1366	Evilgus	0.0041987
Harebrain12	26	Harris1	200	Mustang899	99	Deacona	0.1344	Mustang899	0.0041878
Ibanexxxx	24	Bobster	192	Chillax7	98	Mrjeffreys	0.1337	Phillipjones	0.0041877
T3l3c4st3r	23	Ianburton	172	Lydia1	92	Tscore	0.1335	Harebrain12	0.0041442
DrRock	21	Chaveyjax	163	Bobster	92	Iceman	0.1319	Normanbates	0.0040345
Chaveyjax	20	Tivertom	160	90082	90	Kingkong	0.1291	Yandt	0.0032122
Tunf	19	Tunf	159	Tivertom	71	Zola21	0.1288	Windbag	0.0031122
Mailman	18	Weareborg	156	Clarky	69	Nikkisixx	0.1265	Torres	0.0021265
Sh4rker	18	chilla	129	venus12	68	Mckeith	0.1213	Whatssup	0.0021092

High Page Rank
but low Centrality

TERADATA

Rarely contributes to discussions

User	Activity	Subject
Sh4rker	View	House prices rocketing in some areas
Sh4rker	View	We need to all stop this now!
Sh4rker	View	Smoking outside open windows
Sh4rker	View	Project management best practices attached
Sh4rker	View	Please shut the doors of meeting rooms
Sh4rker	View	Measuring success
Sh4rker	View	My thoughts on the organisation changes
Sh4rker	View	What's the best way to contact IT support nowadays
Sh4rker	View	Unacceptable meeting bookings
Sh4rker	View	What's the best way to contact IT support nowadays
Sh4rker	View	My thoughts on the organisation changes
Sh4rker	View	Join the group masters group now
Sh4rker	View	Summer family fun day
Sh4rker	View	What ad to watch
Sh4rker	View	I need a new pc now!!!!
Sh4rker	View	Iphone 5s or 5c?
Sh4rker	View	New health and safety policy
Sh4rker	View	Website blocked
Sh4rker	View	Antivirus best practive
Sh4rker	View	Global survey – please complete

High number of views on a large range of topics but low number of posts

Compare Users by simple and Graph metrics

Creators	July 13	Views	July 13	Viewers	July 13	Centrality	July13	PageRank	July13
plumbago	85	Daveyboy	1483	Sh4rker	500	Smithy	0.2833	Sh4rker	0.007460
Vazquex9999	69	Ibanexxx	812	Lospollos	158	Tunf	0.2813	Nicky	0.004323
Squire45	49	Liza8	664	Frankenfurt	151	Harris1	0.2515	Tommo14	0.0042112
Chillax7	49	Frankenfurt	557	Ibanexxx	147	Ibanexxx	0.2247	Mrjeffreys	0.0042111
Daveyboy	49	Gill	515	jetsonpete	132	Jetsonpete	0.2238	Jetsonpete	0.0041987
jetsonpete	47	Smithy	380	Daveyboy	128	Bobster	0.2076	Dougcutting	0.0041878
Ibanexxx	42	Baddevotions	378	Jessie	127	Martinez	0.1903	Ffdsa	0.0041877
Baddevotions	38	Jetsonpete	330	Bluerock	121	Massy69	0.1785	Lydia1	0.0041442
mumford	32	63danny	288	Floyd911	118	Meerkat	0.1447	Floyd911	0.0040345
Bobster	31	Martinez	282	Evilgus	115	Langster	0.1440	90082	0.007460
Harris1	30	Johnny	270	Heisenburg	104	Alicewhitey	0.1438	Adamandeveit	0.0042112
Canil	29	Jeff.master	214	Baddevotions	103	Sarahc	0.1403	Squire45	0.0042111
Dotchin	27	maclad	210	Harebrain12	101	Neilneil	0.1366	Evilgus	0.0041987
Harebrain12	26	Harris1	200	[illegible]	[illegible]	[illegible]	0.1344	Mustang899	0.0041878
Ibanexxxx	24	Bobster	192	[illegible]	[illegible]	[illegible]	0.1337	Phillipjones	0.0041877
T3l3c4st3r	23	Ianburton	172	[illegible]	[illegible]	[illegible]	0.1335	Harebrain12	0.0041442
DrRock	21	Chaveyjax	163	Bobster	92	Iceman	0.1319	Normanbates	0.0040345
Chaveyjax	20	Tivertom	160	90082	90	Kingkong	0.1291	Yandt	0.0032122
Tunf	19	Tunf	159	Tivertom	71	Zola21	0.1288	Windbag	0.0031122
Mailman	18	Weareborg	156	Clarky	69	Nikkisixx	0.1265	Torres	0.0021265
Sh4rker	18	chilla	129	venus12	68	Mckeith	0.1213	Whatssup	0.0021092

Similar position across all metrics

TERADATA

A balanced usage

User	Activity	Subject
Jetsonpete	View	Global survey – please complete
Jetsonpete	View	Buy to let rental yields
Jetsonpete	View	Antivirus best practive
Jetsonpete	View	Website blocked
Jetsonpete	View	Please shut the doors of meeting rooms
Jetsonpete	View	Measuring success
Jetsonpete	Create	New version of IE in test phase
Jetsonpete	View	What's the best way to contact IT support nowadays
Jetsonpete	Create	Does anyone have a mini-usb cable I can borrow?
Jetsonpete	View	What's the best way to contact IT support nowadays
Jetsonpete	View	My thoughts on the organisation changes
Jetsonpete	View	Lampard vs Gerard?
Jetsonpete	View	Summer family fun day
Jetsonpete	View	What ad to watch
Jetsonpete	View	New health and safety policy
Jetsonpete	View	Iphone 5s or 5c?
Jetsonpete	Create	Cake in the kitchen this afternoon
Jetsonpete	Create	PC refresh, timetable
Jetsonpete	Create	Reg no A442 WER – you've left your lights on
Jetsonpete	Create	Best place to go all inclusive

High number of posts and views on a large range of topics

TERADATA

User Classification

Fanboys
- Lots of activity but on limited topics

Preachers
- Few Views of other peoples threads but many creates

Lurkers
- Many Views, Few Creates

Collaborators
- Balanced activity across views, creates and topics

Kmeans Clustering

- Use Kmeans to assign users to clusters

Kmeans algorithm clusters users according to the input data – in this case has identified slightly different groupings

Simple SQL based syntax to call algorithm

```
SELECT *
  FROM kmeans(
      ON (SELECT 1)
      PARTITION BY 1
      MAXITERNUM(15)
      NUMBERK('4')
      INPUTTABLE('')
      OUTPUTTABLE(''));
```

```
-1  nopostjoe
-1  beccy
-1  stewartjohns
-1  groovydad2
-1  grownup1

0  Daveyboy
0  PS3only
0  bobthewelder
0  type3
0  masaiplan

1  grahamcole
1  kevild
1  tinatweaks
1  jamiem
1  Claireswy

2  skyler
2  jetsonpete
2  hanshi
2  wolfman
2  c3po

3  Harris1
3  templeton
3  Mynameis6
3  mustardseed
3  clarissa
```

Lurkers
views no posts

Fanboys
most posts on single topic

Potentials collaborators
with low activity

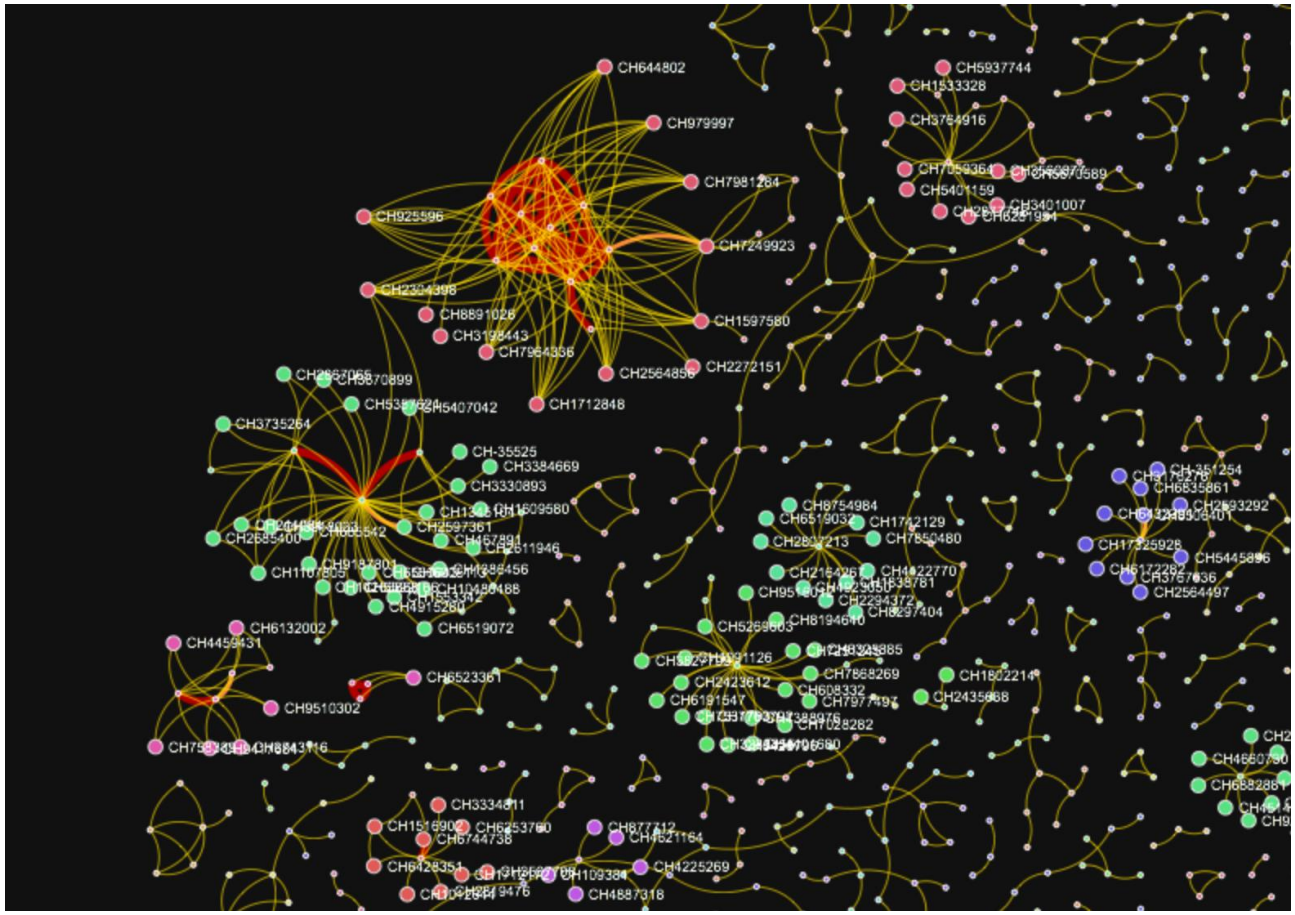
Collaborators
balanced activity

Preachers
posts but few views

TERADATA

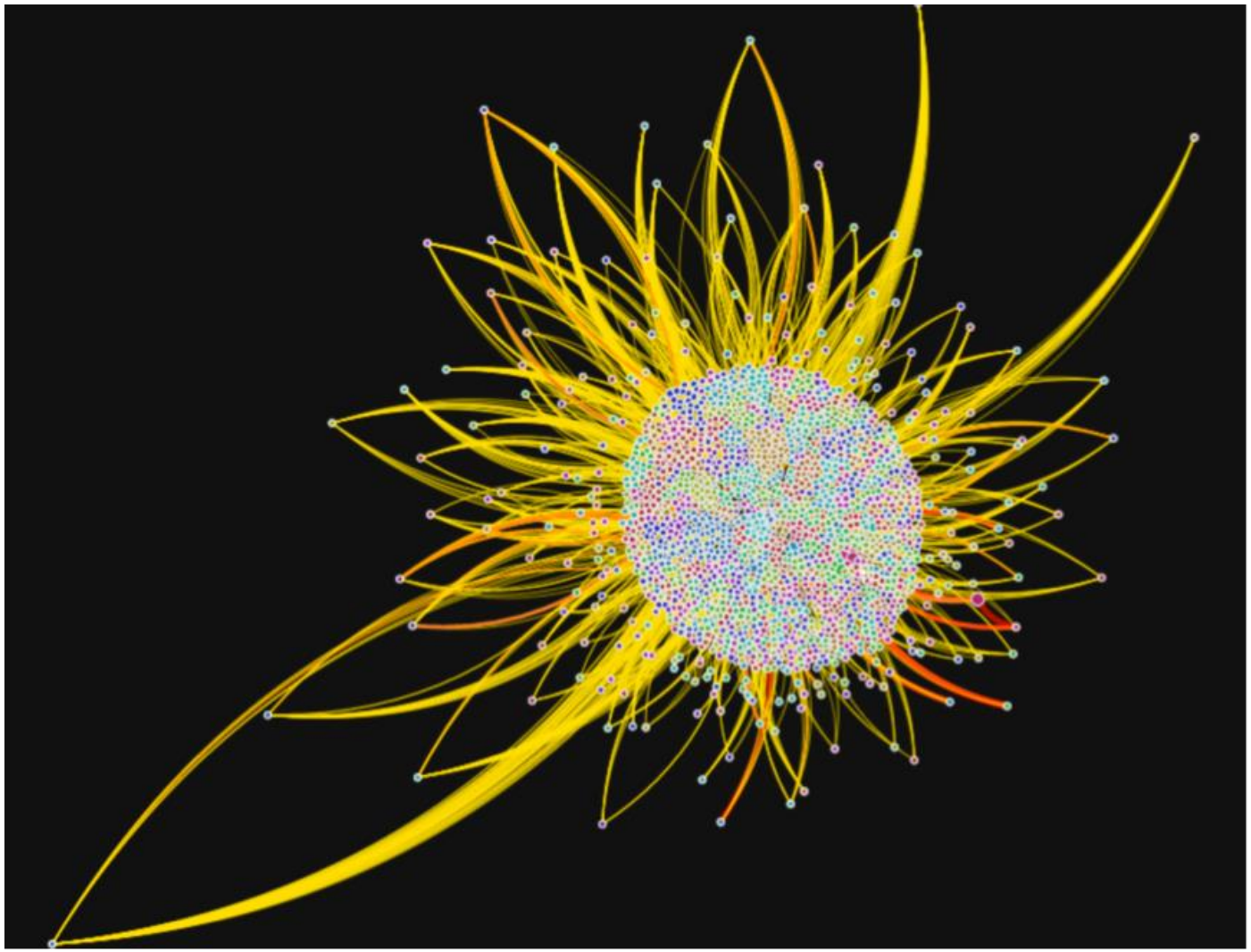
Interactive Graph Visual Demo

- Graphs can be useful for detecting Fraud



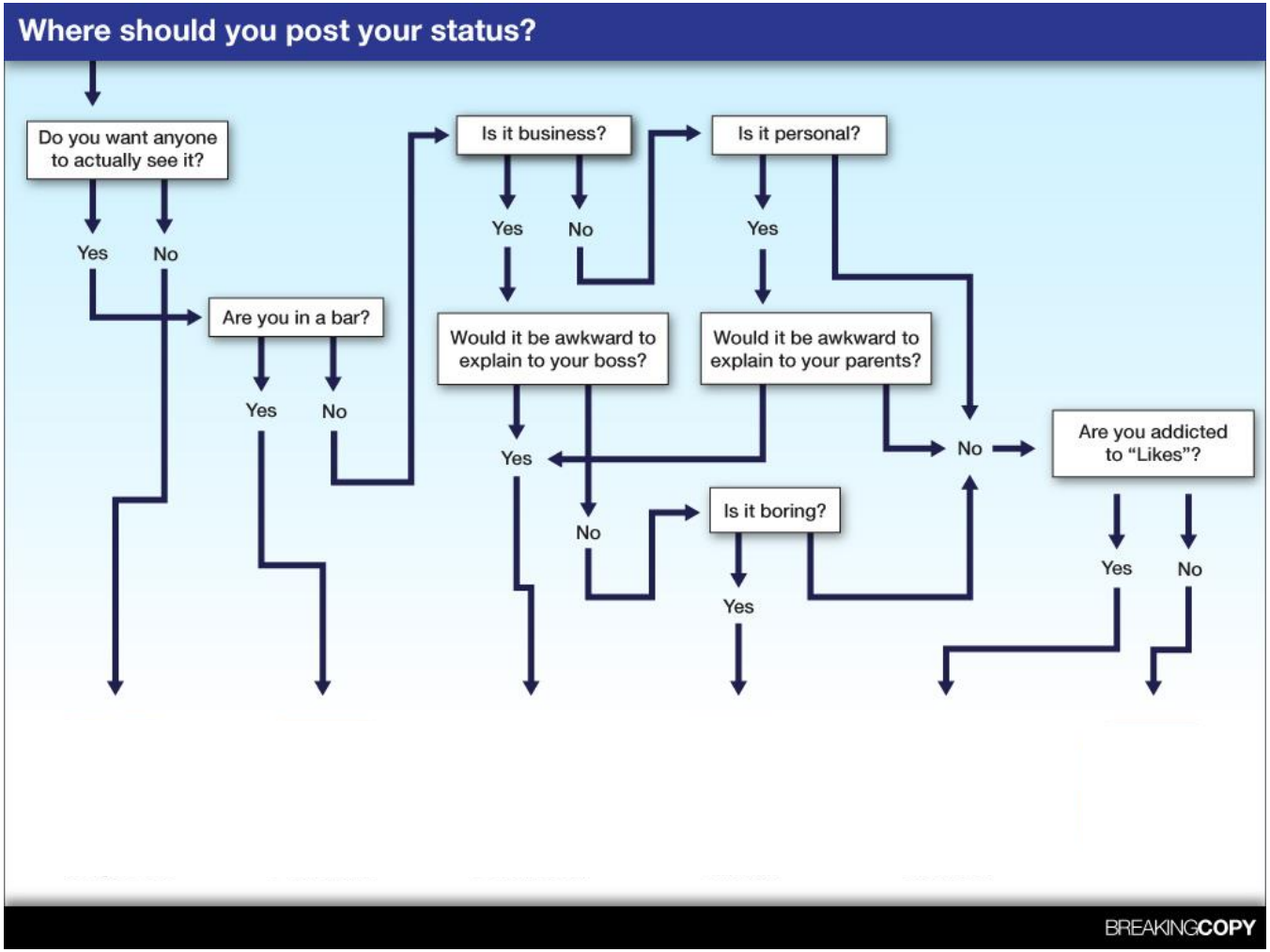
TERADATA

Social data



TERADATA

Social Media



TERADATA

Finding the relevant data

Usea supervised machine learning algorithm for text classification. This takes a pre-classified set of training data and builds a model which can be used to classify other documents

	tweetid	tweettest	tweetclass
1	339837311149826048L	Could not understand a word what the bloke from orange even said to me if he's ...	Telco
2	339834830630035457L	Orange wednesdays anyone?	Telco
3	339782145742630912L	3 people have asked me to do them an orange wednesday today! :/	Telco
4	339751853900312577L	EE sucks! No reception ever what is the point? Orange used to work just fine. Tim...	Telco
5	339723148549292032L	Can anyone on orange sponsor mans a code? Would be much appreciated	Telco
6	339721314904469504L	France Telecom to become Orange on 1 July 2013 http://t.co/LE59AVnUof	Telco
7	339484624000520192L	u201c@poppyparkerx: Orange don't event text me these daysu201d	Telco
8	339442602183106561L	RT @cluedont: I sometimes wonder whether I'd get a better phone signal from an ...	Telco
9	3394372854		
10	3393765517		

	tweetid	tweettest	tweetclass
1	339814269661093888L	#69FactsAboutMe I've got blue green grey eyes with the orange iris but the green...	Not_Telco
2	339062282623537153L	"#GeneralDisturbance at 5570 N ORANGE BLOSSOM TL, Orlando, FL. #orlpol"	Not_Telco
3	339754286089125888L	#guinness #stout has fewer #calories than skim milk or orange juice #vscocam #...	Not_Telco
4	339782699373953026L	#IfiwenttoCO is have some really ugly school colours... No offence but blue and o...	Not_Telco
5	339770030789107712L	#nails #cute #summer #colors #hearts #pink #orange #black #and #white http://...	Not_Telco
6	340467751929905153L	#np blood orange - champagne coast	Not_Telco

TERADATA

Sentiment

Expanding Domain Sentiment Lexicon through Double Propagation

Guang Qiu[*] [1], Bing Liu[**], Jiajun Bu[*] and Chun Chen[*]

[*] College of Computer Science
Zhejiang University
{qiuguang, bjj, chenc}@zju.edu.cn

[**] Department of Computer Science
University of Illinois at Chicago
liub@cs.uic.edu

Abstract

In most sentiment analysis appli
ment lexicon plays a key role. Ho
not impossible, to collect and ma
sentiment lexicon for all applica
cause different words may be us

Generating Domain-Specific Clues using News Corpus for Sentiment Classification

Youngho Kim[§], Yoonjung Choi[+], Sung-Hyon Myaeng[*]

[§]Department of Computer Science, University of Massachusetts Amherst, 140 Governors Drive Amherst, MA, US
[+*]KAIST, 335 Gwahak ro Yuseong gu, Daejeon, South Korea

[§]yhkim@cs.umass.edu, [+]choiyj35@kaist.ac.kr, [*]myaeng@kaist.ac.kr

ain or used together with a specific topic such
in "the plot of Harry Potter is unpredictable".
or development of a domain-specific lexicon
extual information, which would be more

Finding Domain Specific Polar Words for Sentiment Classification

Mehrbod Sharifi
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
msharifi@cs.cmu.edu

William Cohen

Adapting a Polarity Lexicon using Integer Linear Programming for Domain-Specific Sentiment Classification

Yejin Choi and Claire Cardie
Department of Computer Science
Cornell University
Ithaca, NY 14853

A Big Data Approach?

- Access to hundreds of thousands/millions of texts containing consumer sentiment via social media
- Standard “Bag of words” classifiers are not accurate with small amounts of text (e.g. 140 characters)
- What if the data itself can tell us the sentiment

A Big Data Approach?

Use a text tagger on the text that contains emoticons

tweettext	tag
Congratulations orange team :) well done	POS
@keep_orange 17 :) wie alt bist du eigentlich? ich schau grad transformers :) x	POS
Kahapon Blue ngayon naman Red Orange :) #sky @ Camella, Bulakan Bulacan http://t.co/NdBPmODXdv	POS
Every ginger i know is on orange hahahaha, @CamTunstallMuil answer your phonee! :-)	POS
nge right now :(	NEG
	NEG
nja_Orange :D	POS
all about Subic escapade, I guess. :)))) *cto* http://t.co/kL6NsPerOU	POS
	POS
Pourquoi elle est orange sur ses photos ? :(	NEG
Oke, good Charol :) RT @Charols_CnB: @KreasiBogasari iya min itu memang orange marble cake.. :p ~(u02d8u25bdu02d8~)(...	POS
"@detikcom: Foto Bocoran Pertama Leica Mini M Menyeruak http://t.co/ATicQhSUtn via @detikinet" mauuu... :-)	POS
RT @miuius: MIUI.us Version 3.5.31 will be delayed until tomorrow(maybe tonight) MIUI for Nexus 4 Coming this next week. ...	POS
I just want some orange juice :(and soup and meds and well yeah to feel better	NEG
basbakani clockwork orange'taki gibi bi koltuga baglayip su videoyu izletmek lazim, bi ihtimal anlar durumu:) http://t.co/iM...	POS
@mitzifabiaaan osge orange!! :) powta wala munang diet please	POS
my friend is on my macbook and she was drinking orange juice, guess what happened.. >:(	NEG
@MrDrumz242 lol #4 with a Orange soda please :(	NEG
RT @BrooksBeau: Everyone bring green white and orange balloons tomorrow for the end of best friends :) and throw them u...	POS
@DunitJNEY_PV Great ^^ go take some rest sweety :) orange juice? wahh i want i want ^^	POS
@AshleyGrimshaw_ me n @VAntcliffe have made an Indian orange today :) hehe	POS
@orange_juls awesome!!! :)	POS
u201c@Narcoticas: I GOT GOT AN ORANGE AND IT WAS BLACK http://t.co/9ArxMM4aYJu201d there must've been food stam...	POS
"@omozayy: Someone stole my miranda from my fridge at my house party :(I was mad upset" strawbs or orange? cah strawb...	NEG

A Big Data Approach?

tweettext	Bag	Model
I chose a color that was supposed to be a soft peach it came out highlighter orange .. I'm vexed	POS	NEG
@Lsd_no9 @adean_11 I'm gunna turn up like David Dickinson on satdy hahaha #orange	POS	NEG
i thought mixing cranberry juice n orange juice would be delicious but it's rly not :<	POS	NEG
First thing my little cousins says when she sees me ' wow, you look orange' #thanks	POS	NEG
RT @DarloTownPolice: Cops attended a pub last night report of a gun pointed at a male singing kareoke,it was bright orange with flashing lights on	POS	NEG
The #quails' eggs were too dominated by a strong #piccalili/onion accompaniment; #choc/orange #dessert.	POS	NEG
RT @GemmaAnneStyles: My hair is such a beautiful shade of bleach orange/yellow right now. I feel like a low-rent Hayley from Paramore.	POS	NEG
RT @autocorrects: that one person that everyone likes and you're just like "WHY?"	POS	NEG
@LaurenWilson_PA *I am out on the street doing some community service work, in an orange jumpsuit*	POS	NEG
- soo.. only I would miss my cup and spill orange juice all over the counter and floor.. great, just great	POS	NEG
RT @fl511_central: NEW: Incident in Orange on I-4 west at MM 89, right lane blocked.	POS	NEG
Why can't Orange City have a steady Mexican Restaurant #getittogether #cravingit #ugh	POS	NEG
Nothing worse than sitting next to someone on a bus who's oozin bobby orange	POS	NEG
RT @GemmaAnneStyles: My hair is such a beautiful shade of bleach orange/yellow right now. I feel like a low-rent Hayley from Paramore.	POS	NEG
I'm really glad I brushed my teeth before drinking that orange juice.	POS	NEG
RT @annoyingorange: HEY! RT this if you're watching my brand new episode "Orange's Run" on @cartoonnetwork right NOW! HAHAHA!	POS	NEG
I look like Jim Carrey when he drinks orange juice	POS	NEG
RT @GeminiTerms: Trying to read a #Gemini mind is like trying to thinking of a word that rhymes with orange. You just can't do it!	POS	NEG
When I first dyed my hair, it was pinkish... Then it became orange, then like... peachy, and now it's blonde. I wonder went wrong?	POS	NEG
RT @kumpalicious: "remember in 10th grade when you were orange?""you were like a dark orange"	POS	NEG
i dont really like that they're orange but whatever	POS	NEG
I ate Cheetos last night, showered, and my fingers are still orange. Holy cow.	POS	NEG
RT @lukestaceyy: Still don't see how girls think it's attractive to look like an orange	POS	NEG
Orange are right stitch ups!!!	POS	NEG
When girls think they're hot.. like no you've got an orange face and a shelf butt.	POS	NEG
@EE Thanks for trying Dept you put me thru to promised to call me back today but	POS	NEG
You will never see me get excited about Orange Orange when I visit dickenson/wise	POS	NEG

The Modeled approach can be more accurate in predicting domain specific sentiment

TERADATA

We're Hiring

Hadoop Developer Consultant-158116

Description

Position Overview

Teradata is hiring Software Engineers with expertise in Big data and Apache
located in UK, Belgium and Netherlands.

The ideal candidate must be a highly energetic self-starter and with the abili
and run Big Data PoC's for our customers. Depending on experience, the car

Job Specification

Successful candidates will -

- Engage with Teradata Account tea
  requirements;
- Shape and influence customer req
- Assist in qualifying requirements a
  determine whether Hadoop is a go
- Design, plan and execute on-site/c
- Configure and use the Horton Had
  and MapReduce procedural progra
- Partner with the Hadoop administ
  administrate the Hadoop environm
- Post-POC-execution, document an

Data Scientist-159611

Description

About Teradata

Teradata Corporation (NYSE: TDC) is the world's largest company focused on raising intelligence through data warehousing and enterprise analytics. Teradata is in more than 60 countries and on the web at Teradata.com.

Position Overview

The North Europe, Turkey and South Asia Industry Consultant group is looking for Data Scientists that will work with teams to apply Big Data solutions to real-world business situations deriving value for existing customers and proof-of-concept (PoC) projects across the region.

Team Leader Big Data Analytics (m/f)-159672

Description

Teradata helps companies get more value from data than any other company. Our big data analytic solutions, integrated marketing applications, and team of experts can help your company gain a sustainable competitive advantage with data. Teradata helps organizations leverage all of their data so they can know more about their customers and business and do more of what's really important. With more than 10,000 professionals in 43 countries, Teradata serves top companies across consumer goods, financial services, healthcare, automotive, communications, travel, hospitality, and more. A future-focused company, Teradata is recognized by media and industry analysts for technological excellence, sustainability, ethics, and business value. Visit teradata.com.

Currently we are looking for a senior consulting professional for the position of

Team Leader Big Data Analytics (m/f)

Location: Munich, Frankfurt or Düsseldorf

Christopher.hillman@teradata.com
@chillax7

TERADATA

TERADATA

TERADATA ASTER

aprimo
A Teradata Company

Christopher.hillman@teradata.com
@chillax7

TERADATA