

Text analytics *and* new visualization techniques

Richard Brath

rbrath@unchartedsoftware.com

brathr@lsbu.ac.uk

[@rkbrath](#)

richardbrath.wordpress.com

Scott Langevin

slangevin@unchartedsoftware.com

blog.unchartedsoftware.com

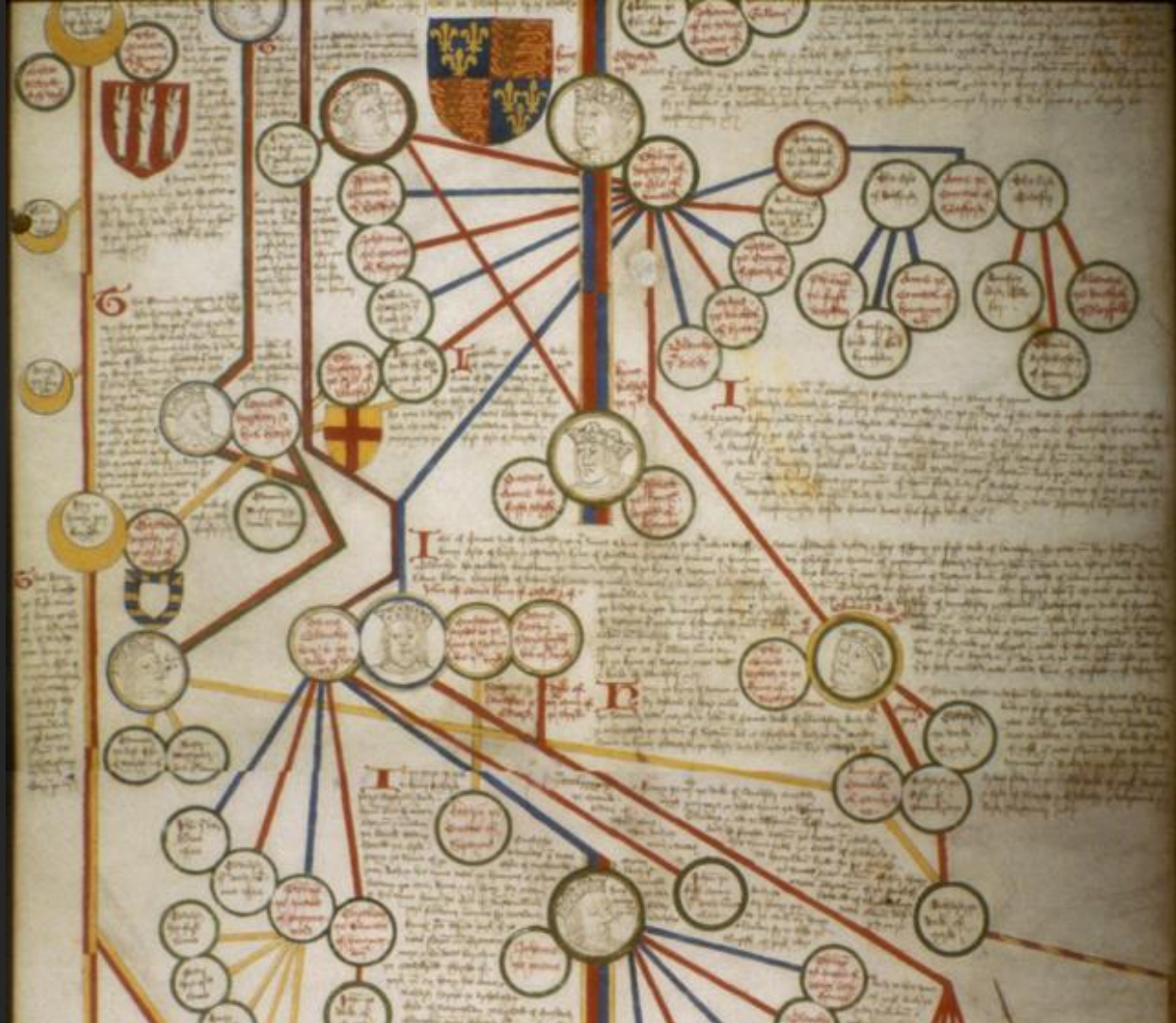
[@slangevi](#)



1. THE 500 YEAR CONSPIRACY

against text visualization

Medieval text and visualization



Genealogical tree from late 1400's. Note graph nodes use of image (people, shield) or text, where text may be black, red or start with a red initial. The nodes can vary in size, color, or shape (circle, crescent, shield). Textual commentary is intertwined throughout.

via [Bodleian Library](#), University of Oxford., Genealogy of the kings of England to Richard III. Chronicle of the Percy family to 1485. MS. Bodl. Rolls 5, view 35.

1439 Johannes Gutenberg changes everything



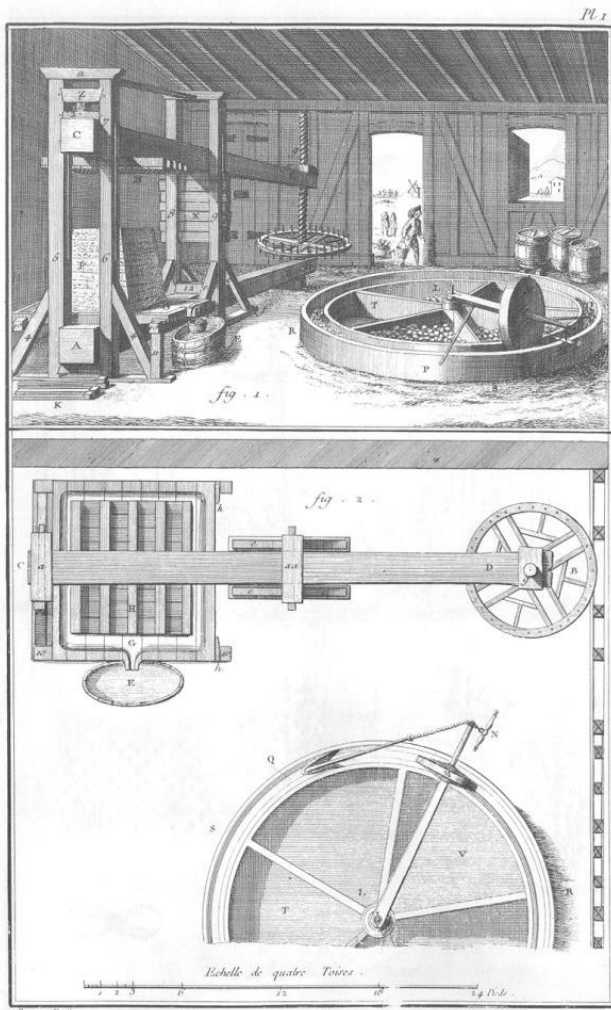
1573: Image separates from text.
from William Bullein's A Dialogue... Against the
Fever Pestilence. Author photo from Bodleian exhibition
"Shakespeare's Dead".

1751 Enlightenment complete separation

Diderot's Encyclopedia has great illustrations of various occupations – all neatly labeled, but the viewer has to cross-reference the text to understand.

The *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, par une Société de Gens de lettres was published under the direction of Diderot and d'Alembert, with 17 volumes of text and 11 volumes of plates between 1751 and 1772. Containing 74,000 articles written by more than 130 contributors, the *Encyclopédie* was a massive reference work for the arts and sciences, as well as a *machine de guerre* which served to propagate the ideas of the French Enlightenment. The impact of the *Encyclopédie* was enormous. Through its attempt to classify learning and to open all domains of human activity to its readers, the *Encyclopédie* gave expression to many of the most important intellectual and social developments of its time.

Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers, etc., eds. Denis Diderot and Jean le Rond d'Alembert. University of Chicago: ARTFL Encyclopédie Project (Spring 2016 Edition). Robert Morrissey and Glenn Roe (eds). <http://encyclopedie.uchicago.edu/>.



Benard Peul

L'Economie Rustique,
Presoir à Cidre.

Pl. 1.

ET ECONOMIE RUSTIQUE. 111

PLANCHE III.

- Fig. 1. Plan & profil de l'un des coffres du pressoir. P P, chantier. r r, brebis. y, doffier. q, le mulet. y, y, y, y, pièces de maitre. Z, coins. D, mouton. E E, coins ou pouilleux. u u, écrou. C D, vis. A B, grande roue.
2. Coupe suivant la longueur d'un des coffres du pressoir. L L, faux chantier. 13, 13, jumelles. e e, chaînes. y, doffier. r r, brebis. q, le mulet. 1, 2, 3, 4, 5, 6, 7, 8, 9, pièces de maitre. Z, coins. D, mouton. E E, coins ou pouilleux. p, p, p, p, appuis du doffier. 10, mouleau. G G, planches à couleaux. x, x, x, x, chevrons. u u, écrou. C D, vis. A B, grande roue. M, bouquets ou pieds-taux de pierre. F, le marc.

PRESOIR A CIDRE. PLANCHE I^{re}.

- Fig. 1. Vue perspective & plan du pressoir. La vignette représente l'endroit où le pressoir & la pile font établis. A B, la brebis. C D, le mouton. 1, 6, 7, 8, 9, les jumelles. 4, 4, 4, 4, contrevens. Y, Z, 2, 12, entretoises. a, b, chapeau. K X, les clés. g, la vis. E, le batlong. F, marc empilé sur la maitre ou l'emoy. G, la maitre ou l'emoy. 10, 10, pièces qui supportent les couches. H, le hec. R S Q, auge circulaire de la pile. Q, le robot. T U V, écrautes. 11, purgier ou grand moulin servant à mettre les sucres quand ils sont en forme, pour les purger de leur syrop superflu & les terrier. P, palonnier. V P, conducteur du cheval ou guide.
2. Plan du pressoir & de la pile. Les mêmes lettres désignent les mêmes parties.

PLANCHE II.

- Fig. 3. Profil & détail du pressoir à cidre. Élévation géométrale du pressoir vu de face. Les mêmes lettres désignent aussi les mêmes parties.
4. Élévation des jumelles qui embrassent le gros bout du mouton & de la brebis.
5. Élévation des deux jumelles qui sont placées vers le milieu du mouton & de la brebis, & qui servent à relever le mouton.
6. Partie inférieure de la vis qui entre dans la brebis.
7. Plan & profil d'une des clés.

INDIGOTERIE ET MANIOC.

LE haut de la Planche ou la vignette représente la vue d'une indigoterie. A, réservoir d'eau claire. B, la trempoire. C, la batterie. D, le repoiroir qu'on nomme aussi diabolotin. E E, robinets d'où la teinture d'une cuve pallé dans la cuve qui est au-dessous. E F, trous que l'on débouche successivement, pour verser l'eau claire de la batterie. Lorsque la feuille bleue s'est précipitée au fond. G, indigot dont on a rempli des sacs de toile en forme de chauffes pour le faire évouer. H, hangard ouvert & à claire voie sous lequel on met l'indigot dans des caissons, pour avoir et le faire sécher à l'ombre. I, negre qui porte la plante dans la trempoire. K K, negres qui agitent continuellement la teinture de la batterie avec des lances percés & attachés à de longues perches. L, planes d'indigot. M, maison du maître de l'habitation. N, campagne fermée.

- Fig. 1. o o, caissons de bois élevés sur des treteux, servant à faire sécher l'indigot à l'ombre sous le hangard de la vignette.
2. P, contens courbé en forme de serpette, pour couper l'indigot par pié.
3. Q, talle d'argent bien polie, servant à examiner la formation du grain dans la teinture de la batterie.
4. Presse à manioc. A, tronc d'arbre percé en-travers. B, branche fourchue disposée en bras de levier &

chargée de grosses pierres. C, sacs d'écorce d'arbre remplis de la rapure du manioc. D, bouts de planche servant à presser les sacs également. E, cour ou coupe de camballe recevant le suc du manioc dont on fait la mouchoue.

Fig. 5. Manière d'imprimer le suc du manioc à la façon des Carabes. G, couleuvre ou espèce de panier d'un tissu liège & flexible, rempli de rapure de manioc. H, poids attaché au bas de la couleuvre; qui le contraind de s'allonger en diminuant la grosseur; ce qui suffit pour exprimer le suc de la rapure.

SUCRERIE ET AFFINAGE des Sucres.

PLANCHE I^{re}.

- L' vignette représente la vue d'une habitation. 1, maison du maître & ses dépendances. 2, 2, 2, partie des cases à negres formant une ou plusieurs rues, suivant le nombre & l'emplacement. 3, 3, 3, partie de favanne ou pîtrage. 4, 4, lierre ou forte haie qui sépare les plantations de cannes. 5, 5, 5, partie de pieces plantées en cannes à sucre à mi-côte & en plat-pays. 6, moulin à eau. 7, fucrière avec sa cheminée, & son hangard pour les fourneaux. 8, gouttière qui conduit l'eau du canal sur la roue du moulin. 9, décharge de l'eau du moulin. 10, une des cases à bagalles ou cannes écrautes. 11, purgier ou grand moulin servant à mettre les sucres quand ils sont en forme, pour les purger de leur syrop superflu & les terrier. 13, hauteurs entre lesquelles font les plantations de manioc. Les bananiers de l'habitation à vivre. 14, morne; c'est ainsi qu'on nomme aux îles Antilles les montagnes qui paroissent détachées des autres.
- Fig. 1. Coupe verticale d'une étuve à mettre sécher les pains de sucre terrez. A, comble de l'étuve. B, murs de l'étuve. C, porte. D, coffre de fer servant de fourneau. E, bouches du foyer & du cendrier. F, rayons ou tablettes en grillage, sur lesquelles on range les pains de sucre. G, plancher couvert de cinq à six pouces de maçonnerie. H, trape que l'on ouvre pour laisser aller l'humidité qui s'éleve des pains de sucre, & qui s'échappe au-dehors par les conduits i, i, pratiqués sous le larmier.
2. K, canne à sucre. L, feuille détachée sur les bords. M, fleche ou fleur de la canne portant la graine. N, partie inférieure de canne avec racine.
3. O, serpe pour farcler & couper les cannes.
4. P, houe à fouiller la terre.
5. Q, pelle de fer pour le même usage & ramasser le sucre pilé dans le canot.
6. R, pelle de fer servant de levier.
7. S, canot avec des pilons, pour mêler le sucres en poudre, & le fouler dans les futailes.

PLANCHE II.

Fig. 1. Moulin mû par des animaux. AA, châssis de charpente très-mollet. B, table du moulin, communément faite d'un seul bloc creusé & revêtu de plomb. C, C, C, trois roles couverts chacun d'un tambour ou cylindre de métal, & traversés d'un axe de fer coulé, dont l'extrémité inférieure est garnie d'un pivot portant sur un crapaudin. D, D, D, D, ouvertures faites à la table pour pouvoir charger & réparer les pivots & les crapaudines. E, E, entailles aux deux ouvertures des côtés, servant à chasser des coins de bois, pour ferret & rapprocher les tambours. F, fentes ouvertes sur les moifles, avec des coins pour ferret les pivots supérieurs. G, G, héritons dont les roles font couronnés, & qui engrenent les uns dans les autres, & que tournent les tambours en sens contraires. H, axe ou arbre prolongé du principal role. I, demofelle,

1920 Same for charts

more than would be required for the tabulated figures, and if there is any serious limitation on space, Fig. 97 could be much reduced in size without detracting from its clearness.

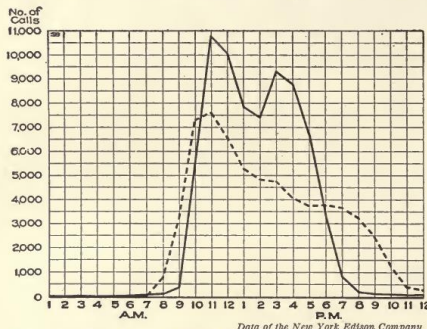


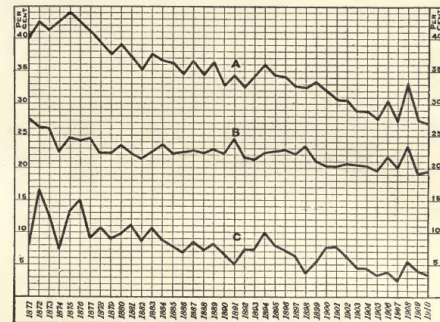
Fig. 97. Number of Telephone Messages Each Hour for One Day in a Business District in New York City as Compared with a Residence District

Solid line, the "Broad" exchange—a typical business exchange
Dotted line, the "Riverside" exchange—a typical residence exchange
Note the great number of business calls after mail has been opened in the morning and after return from lunch. In the residence district there is much morning shopping by telephone

Though the colored inks assisted in catching the eye of the reader, the chart with curves designated by letters is usually sufficient for all practical purposes whenever the curves do not cross over each other in such manner as to be confusing. As mentioned elsewhere, a certain slope of a curve plotted on rectangular co-ordinate paper does not in itself indicate a greater or less amount of increase or decrease than holds true for some other curve having a different slope. The slope of a curve plotted on paper with ordinary co-ordinate ruling depends largely on whether the data of the curve are in large figures, so as to bring the curve near the top of the chart, or in small figures, bringing the curve near the bottom of the chart. In Fig. 98 the fact that curve A slopes more than curve C is due to the fact that curve A is placed higher on the vertical scale of the chart than curve C. A little study will show that the reduction in mortality portrayed by curve A is much less on a percentage basis than that depicted by curve C, yet curve A has the steeper slope. The slope of these two

curves can best be compared if a pencil line is drawn in such manner that the peaks above the pencil line are approximately equal to the valleys below the pencil line for each curve.

Fig. 99 contains some interesting information. Though the chart proves fairly well the close dependence of the price of cast-iron pipe upon the price of the pig iron from which it is made, the chart is nevertheless misleading in that the first glance would indicate a much greater fluctuation in the price of pig iron than actually occurred.



From the 1910 Annual Report of the Health Department, City of Boston

Fig. 98. Deaths in Boston of Children under Five Years of Age, under One Year, and from Five of the Principal Infectious Diseases, Expressed as a Percentage of the Total Mortality

Curve A. Deaths of children under five years of age as a percentage of the total mortality
Curve B. Deaths of children under one year as a percentage of the total mortality
Curve C. Deaths from Diphtheria, Scarlet Fever, Measles, Typhoid Fever and Smallpox as a percentage of the total mortality
Note that Curve A shows a much steeper slope than Curve C, yet Curve C drops in 1910 to less than half the figure for 1871. Curves plotted by rectangular co-ordinates should not be compared by the slope of the different curve lines

The reader is apt to overlook the fact that the vertical scale of the chart does not extend below \$11 per ton. He is quite likely to think that the price of pig iron had all the rapid fluctuations which would be indicated by the changing vertical distances between the pig-iron curve and the bottom line of the chart itself. The amount of fluctuation would look much less if the chart extended to the zero line of the vertical scale.

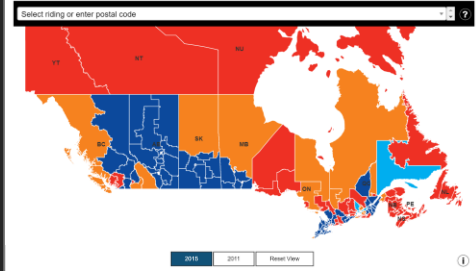
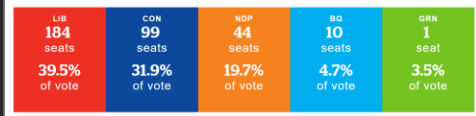
1919 book explaining charts. Labels and notes are pushed to the periphery of the chart. (Willard C. Brinton, *Graphic Methods for Presenting Facts*, 1919)

Federal election results 2015

Justin Trudeau's Liberals to form majority government

Justin Trudeau's Liberals won a decisive majority government, capturing 184 seats. The Conservatives, who were seeking their fourth consecutive mandate, will now form the Official Opposition. Find detailed ridings results in our [map](#) - see where each party gained and lost support and view the demographic split by party and riding. A total of 85 women were sent to House, see how the vote splits by party. And, track the Liberals surge in the polls over the course of the 78-day campaign.

UPDATED Oct 26, 2015, 1:00 AM EST



Jump to a region:
[Vancouver](#) [Calgary](#) [Winnipeg](#) [Toronto](#) [Montreal](#) [Atlantic](#)

Explore the map fullscreen



Bruce Anderson, one of CBC's At Issue panelists, called Justin Trudeau's ascent to prime minister as a "campaign for the ages." Sure, Trudeau is the son of a former PM. But he defied great odds to win.

The Liberals started the campaign with a mere 37 seats in the House of Commons and were polling in third place. But once the ballots were all counted, the Liberals had taken 184 seats.

Trudeau's astounding success also highlights the reversals of fortune for both the Conservatives, who have governed since 2006, and the NDP, who were first in the polls going into this 11-week campaign.

Stephen Harper and the Conservatives won the 2011 election with 166 seats, while the NDP finished with 103 — and, for the first time, status as the Official Opposition. Four years later, the Conservatives have 67 fewer seats while the NDP have dropped 59.

ELECTION 2012 Presidential Race House and Senate Campaign Finance Debates Results

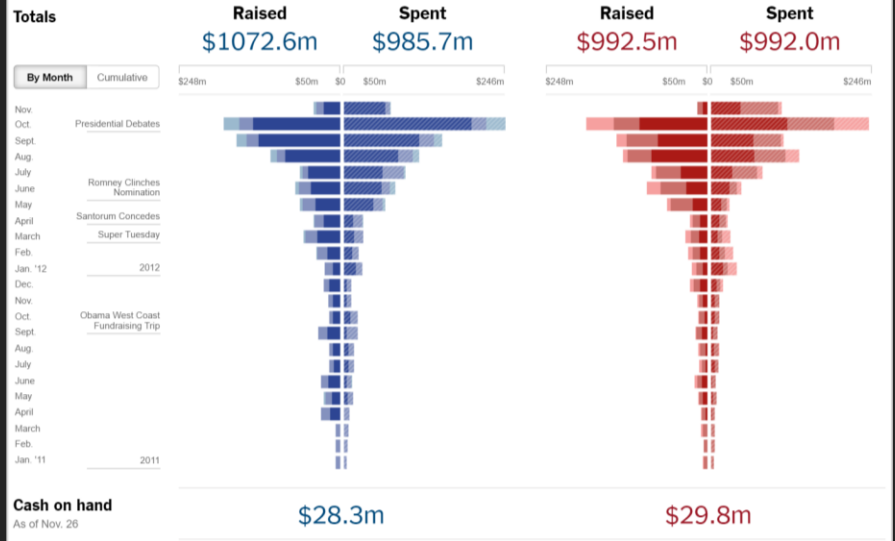
The 2012 Money Race: Compare the Candidates

FACEBOOK TWITTER

Below is a tally of the money raised and spent through September by the presidential candidates, the national party committees and the primary "super PACs" whose sole purpose is to support a candidate. Contribution and spending totals do not include money raised or held by each candidate's "victory fund," a joint fund-raising committee that will distribute funds to the campaigns and party committees. In addition to these committees, nonprofit groups that do not have to file with the Federal Election Commission and other super PACs have spent at least \$65 million more on television advertising, almost all of it against President Obama or in support of Mitt Romney.

Barack Obama +
Democratic Party +
Priorities USA Action Super PAC

Mitt Romney +
Republican Party +
Restore Our Future Super PAC



German election: Merkel vows to win back right-wing voters

3 hours ago Europe

Europe migrant crisis

The German Chancellor Angela Merkel has said she wants to win back the voters who deserted her party for right-wing nationalist AfD.

Mrs Merkel is set for a fourth term in office, despite heavy losses for her party.

She now faces months of coalition talks with several parties in order to form a stable government.

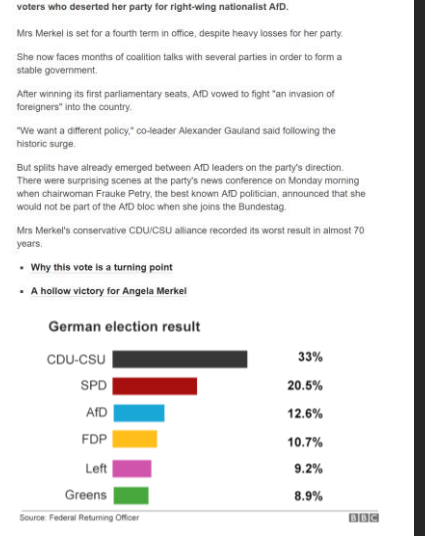
After winning its first parliamentary seats, AfD vowed to fight "an invasion of foreigners" into the country.

"We want a different policy," co-leader Alexander Gauland said following the historic surge.

But splits have already emerged between AfD leaders on the party's direction. There were surprising scenes at the party's news conference on Monday morning when chairwoman Frauke Petry, the best known AfD politician, announced that she would not be part of the AfD bloc when she joins the Bundestag.

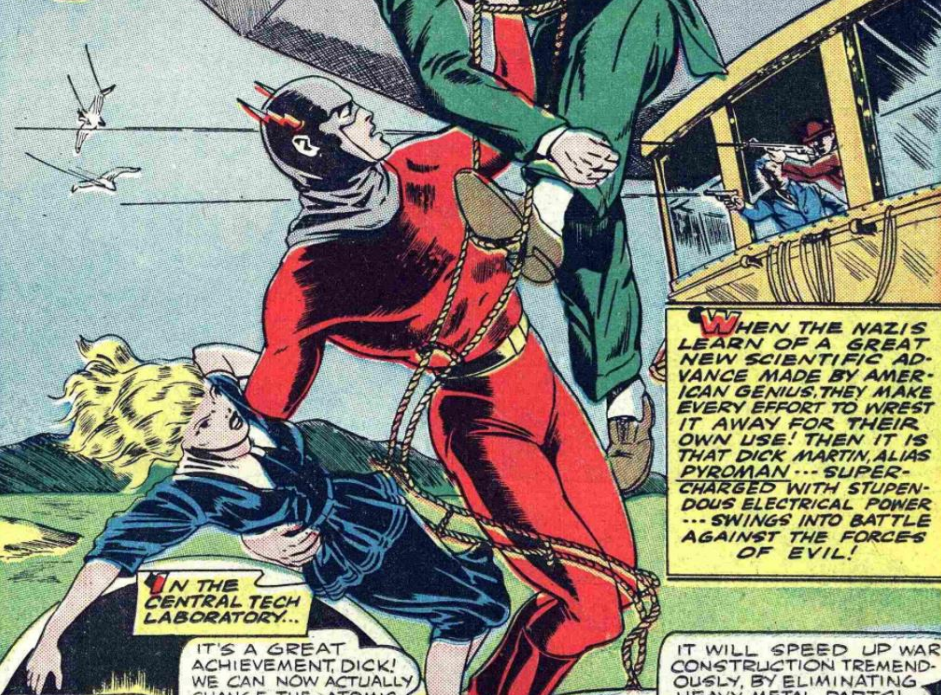
Mrs Merkel's conservative CDU/CSU alliance recorded its worst result in almost 70 years.

- Why this vote is a turning point
- A hollow victory for Angela Merkel



PYROMAN

Text and Imagery working together. Digital Comic Museum. Pyroman, from America's Best Comics; volume 13, 1945. Public Domain.
<https://digitalcomicmuseum.com/preview/index.php?dId=13773&page=2>



WHEN THE NAZIS LEARN OF A GREAT NEW SCIENTIFIC ADVANCE MADE BY AMERICAN GENIUS, THEY MAKE EVERY EFFORT TO WREST IT AWAY FOR THEIR OWN USE! THEN IT IS THAT DICK MARTIN, ALIAS PYROMAN ... SUPER-CHARGED WITH STUPENDOUS ELECTRICAL POWER ... SWINGS INTO BATTLE AGAINST THE FORCES OF EVIL!

IT WILL SPEED UP WAR CONSTRUCTION TREMENDOUSLY, BY ELIMINATING

IN THE CENTRAL TECH LABORATORY...

IT'S A GREAT ACHIEVEMENT, DICK! WE CAN NOW ACTUALLY

OUTSIDE THE LABORATORY...

DER BOSS SAID DIS LABORATORY ISS DEVELOPING AN IMPORTANT WAR SECRET, HEINRICH! IT ISS UP TO US TO FIND OUT MORE ABOUT IT!



JA! WE WILL...WAIT! HERE COME TWO OF DER LABORATORY WORKERS!

SURE...THE NEW RADIO-ACTIVATOR TURNS ANY GAS INTO A SOLID... INSTANTANEOUSLY!

DID YOU HEAR THAT, HEINRICH? DER BOSS WAS RIGHT!

YOU CAN'T DO THIS! I...

WE DO WHATEVER WE WANT! NAZI GERMANY WILL NOT BE STOPPED!

THAT EVENING...

UP WITH YOUR HANDS...QUICK! HANS, GRAB DER RADIO-ACTIVATOR!

WHAT THE...!



HERE'S WHERE I MAKE A TRY AT IT, ANYWAY!

UGH!

I SHOULD KILL YOU...BUT WE NEED YOU TO SHOW US HOW TO WORK DER MACHINE! I...

BOSS! DER POLICE! DEY MUST HAVE HEARD DER FIGHTING!

AT THAT MOMENT... JOYCE CLARK!

VERDAMMT! WE WILL HAVE TO GET AWAY FAST! COME!

DICK! WHAT...

2. Bringing **TEXT** back into **VISUALIZATION**

Plain Dots

- Simple scatterplot
- Macro pattern e.g. crescent shape
- Also color and size set to population

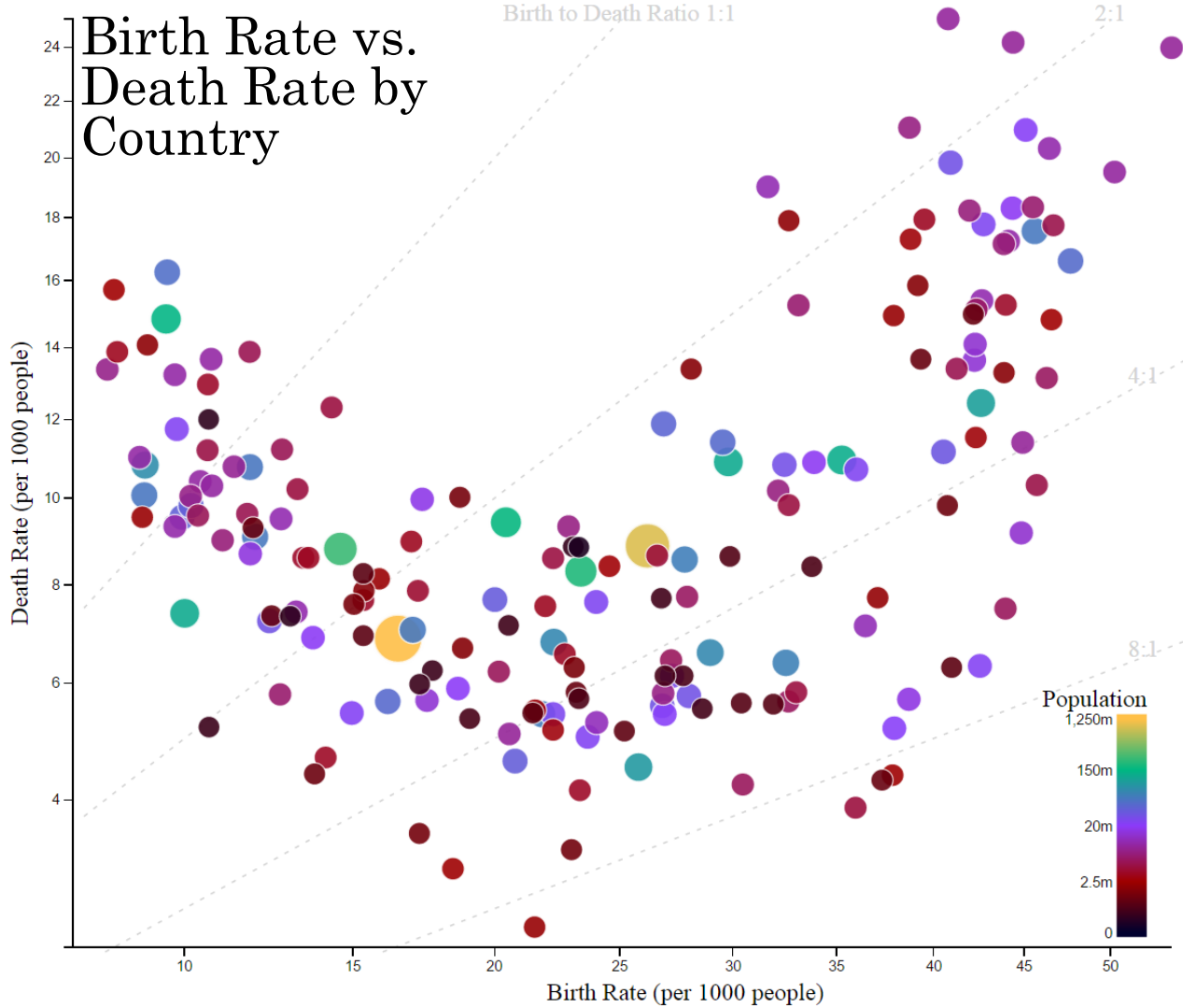


Image: Birth rate vs. Death rate per country, with size and color indicating population; and diagonal lines indicating birth to death ratio. Note visible crescent shape. Image by author.

Alphanumeric Points

- Same scatterplot
- 2-letter ISO codes added
- More information, more insight

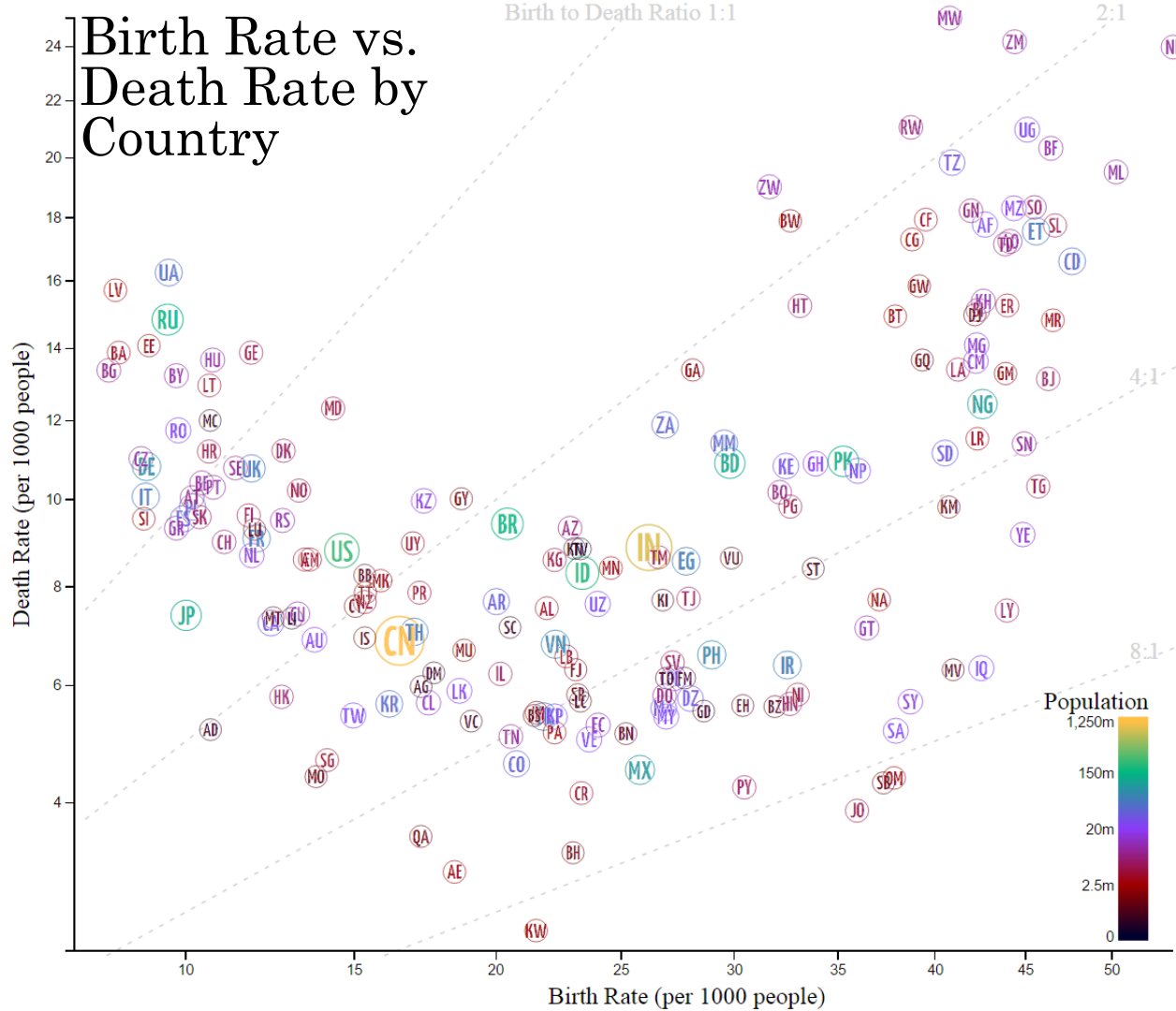


Image: 2 letter country ISO codes act as mnemonics. Note codes upper left: LV, UA, RU, EE – Latvia, Ukraine, Russia, Estonia, etc., all below population replacement rate. Image by author.

Best-Selling Singles more than 10 million copies, sorted by release date

Bar Chart of Text

Better than label and bar: See artist, title and first line.

Simple label has problems: Artist as label doesn't show the song. Song as label is ambiguous (e.g. who's version of *Silent Night*?). Label using artist + song is very long and leaves little space for bar (e.g. see entry for Elton John). First line provides rich context: maybe you don't know the *Wabash Cannonball*, but with the first line you know what it's about. Images by author.

Bing Crosby -- "Silent Night" ~ Silent night, Holy night, All is calm, All is bright, Round your virgin mother and child, Holy infant more than words can say, If I didn't care, would I feel this way? If this isn't	1
The Ink Spots -- "If I Didn't Care" ~ If I didn't care more than words can say, If I didn't care, would I feel this way? If this isn't	2
Bing Crosby -- "White Christmas" ~ I'm dreaming of a white Christmas, Just like the ones we used to know, Where the treetops glisten	3
Roy Acuff -- "Wabash Cannonball" ~ From the great Atlantic ocean to the wide Pacific shore, From the queen of flowing mountain to the	4
Wills Brothers -- "Paper Doll" ~ I'm gonna buy a paper doll that I can call my own, A doll that other fellows cannot steal, And their	5
Gene Autry -- "Rudolph the Red-Nosed Reindeer" ~ You know Dasher and Dancer and Prancer and Vixen, Comet and Cupid and Donner and Blitzen	6
Patti Page -- "Tennessee Waltz" ~ I was dancing with my darling to the Tennessee Waltz, When an old friend I happened to see, I introduced	7
Bill Haley & His Comets -- "Rock Around the Clock" ~ One, two, three o'clock, four o'clock, rock, Five, six, seven o'clock, eight o'clock	8
The Penguins -- "Earth Angel" ~ Earth angel, earth angel, Will you be mine?, My darling dear, Love you all the time, I'm just a fool, You	9
Elvis Presley -- "Hound Dog" ~ Hey baby I'm here to tell you about yourself, You ain't nothin' but a hound dog, scratchin' all the time, I'm	10
Paul Anka -- "Diana" ~ I'm so young and you're so old, This, my darling, I've been told, I don't care just what they say, 'Cause fore and	11
Domenico Modugno -- "Nel blu dipinto di blu (Volare)" ~ Sometimes the world is a valley of heartaches and tears, And in the hustle and bustle of	12
Elvis Presley -- "It's Now or Never" ~ It's now or never, Come hold me tight, Kiss me my darling, Be mine tonight, Tomorrow will be too	13
Kyu Sakamoto -- "Sukiyaki" ~ Let's look up as we walk, so that the tears don't spill, Remembering that spring day, that lonely night.	14
The Beatles -- "I Want to Hold Your Hand" ~ Oh yeah I tell you somethin', I think you'll understand, When I say that somethin', I want	15
The Monkees -- "I'm a Believer" ~ I thought love was only true in fairy tales, Meant for someone else but not for me, Love was out to	16
Procol Harum -- "A Whiter Shade of Pale" ~ We skipped the light fandango, Turned cartwheels 'cross the floor, I was feeling kinda sea	17
Mungo Jerry -- "In the Summertime" ~ In the summertime when the weather is hot, You can stretch right up and touch the sky, When the	18
Middle of the Road -- "Chirpy Chirpy Cheep Cheep" ~ Where's your mamma gone, Little baby bird, Where's your mamma gone, Far far away.	19
Carl Douglas -- "Kung Fu Fighting" ~ Now here it is, one to make you move, Something with a funky kung fu groove, Something that'll make	20
George McCrae -- "Rock Your Baby" ~ Sexy mama, Woman, take me in your arms, Rock your baby, Woman, take me in your arms, Rock your baby	21
Roger Whittaker -- "The Last Farewell" ~ There's a ship lies rigged and ready in the harbor, Tomorrow for old England she sails, Far, far	22
ABBA -- "Fernando" ~ Can you hear the drums Fernando, I remember long ago another starry night like this, In the firelight Fernando,	23
John Travolta & Olivia Newton-John -- "You're the One That I Want" ~ I got chills, They're multiplying, And I'm losing control, Cause you're	24
Gloria Gaynor -- "I Will Survive" ~ At first I was afraid, I was petrified, Kept thinking I could never live without you by my side, I	25
Village People -- "Y.M.C.A." ~ Young man, there's no need to feel down, I said, young man, pick yourself off the ground, I said, young	26
Trio -- "Da Da Da" ~ Aha aha aha, Aha, aha, aha, Was ist los mit dir mein Schatz? Aha, Geht es immer nur bergab? Aha, Geht nur das was	27
Band Aid -- "Do They Know It's Christmas?" ~ It's Christmas time, there's no need to be afraid, At Christmas time, we let in light and	28
USA for Africa -- "We Are the World" ~ There comes a time, When we heed a certain call, When the world must come together as one, The	29
Kaoma -- "Lambada" ~ Chorando se foi quem um dia so me fez chorar, Chorando se foi quem um dia so me fez chorar, Chorando estara, ad	30
Bryan Adams -- "(Everything I Do) I Do It for You" ~ Look into my eyes, You will see, what you mean to me, Search your heart, search your	31
Scorpions -- "Wind of Change" ~ I follow the Moskva, Down to Gorky Park, Listening to the wind of change, An August summer night, Sol	32
Whitney Houston -- "I Will Always Love You" ~ If I should stay, I'll only be in your way, So I'll go, but I know I'll, Think of you every	33
Los del Río -- "Macarena" ~ Dale a tu cuerpo alegría Macarena, Que tu cuerpo es pa' darle alegría y cosa buena, Dale a tu cuerpo alegría	34
Andrea Bocelli & Sarah Brightman -- "Time to Say Goodbye" ~ Quando sono sola, Sogno all'orizzonte, E mancan le parole, Si' lo so che, I	35
Toni Braxton -- "Un-Break My Heart" ~ Don't leave me in all this pain, Don't leave me out in the rain, Come back and bring back my love	36
Elton John -- "Candle in the Wind 1997"/"Something About the Way You Look Tonight" ~ Goodbye Norma Jean, Though I never knew you at	37
Celine Dion -- "My Heart Will Go On" ~ Every night in my dreams, I see you, I feel you, That is how I know you, go on, Far across the	38
Cher -- "Believe" ~ No matter how hard I try, You keep pushing me aside, And I can't break through, There's no talking to you, It's so	39
Punjabi MC -- "Mundian to Bach Ke" ~ It's the Roc in the building, Calib, Rame!, Tarrell in the house, Minian to but the tiri bachin	40
Britney Spears -- "...Baby One More Time" ~ Oh baby, baby, how was I supposed to know, That something wasn't right here, Oh baby, baby	41
Celine Dion & Andrea Bocelli -- "The Prayer" ~ I pray you'll be our eyes and watch us where we go, And help us to be wise in times when	42

Best-Selling Singles more than 10 million copies, sorted by release date

Bar Chart of Text

Plus content analysis:

Is "love" the top word in the top 40?

Bing Crosby -- "Silent Night" ~ Silent night, Holy night, All is calm, All is bright, Round your virgin mother and child, Holy infant the Inks Spots -- "If I Didn't Care" ~ If I didn't care more than words can say, If I didn't care, would I feel this way? If this isn't				
Bing Crosby -- "White Christmas" ~ I'm dreaming of a white Christmas, Just like the ones we used to know, Where the treetops glisten				
Roy Acuff -- "Wabash Cannonball" ~ From the great Atlantic ocean to the wide Pacific shore, From the queen of flowing mountain to the				
Wills Brothers -- "Paper Doll" ~ I'm gonna buy a paper doll that I can call my own, A doll that other fellows cannot steal, And then				
Gene Autry -- "Rudolph the Red-Nosed Reindeer" ~ You know Dasher and Dancer and Prancer and Vixen, Comet and Cupid and Donner and Bl				
Patti Page -- "Tennessee Waltz" ~ I was dancing with my darling to the Tennessee Waltz, When an old friend I happened to see, I intro				
Bill Haley & His Comets -- "Rock Around the Clock" ~ One, two, three o'clock, four o'clock, rock, five, six, seven o'clock, eight o'clock				
The Penguins -- "Earth Angel" ~ Earth angel, earth angel, Will you be mine?, My darling dear, Love you all the time, I'm just a fool,				
Elvis Presley -- "Hound Dog" ~ Hey baby I'm here to tell you about yourself, You ain't nothin' but a hound dog, scratchin' all the ti				
Paul Anka -- "Diana" ~ I'm so young and you're so old, This, my darling, I've been told, I don't care just what they say, 'Cause fore				
Domenico Modugno -- "Nel blu dipinto di blu (Volare)" ~ Sometimes the world is a valley of heartaches and tears, And in the hustle an				
Elvis Presley -- "It's Now or Never" ~ It's now or never, Come hold me tight, Kiss me my darling, Be mine tonight, Tomorrow will be t				
Kyu Sakamoto -- "Sukiyaki" ~ Let's look up as we walk, so that the tears don't spill, Remembering that spring day, that lonely night.				
The Beatles -- "I Want to Hold Your Hand" ~ Oh yeah I tell you somethin', I think you'll understand, When I say that somethin', I wan				
Elvis Presley -- "I'm a Believer" ~ I thought Love was only true in fairy tales, Meant for someone else but not for me, Love was out to				
Procol Harum -- "A Whiter Shade of Pale" ~ We skipped the light fandango, Turned cartwheels 'cross the floor, I was feeling kinda sea				
Mungo Jerry -- "In the Summertime" ~ In the summertime when the weather is hot, You can stretch right up and touch the sky, When the				
Middle of the Road -- "Chirpy Chirpy Cheep Cheep" ~ Where's your mamma gone, Little baby bird, Where's your mamma gone, Far far away.				
Carl Douglas -- "Kung Fu Fighting" ~ Now here it is, one to make you move, Something with a funky kung fu groove, Something that'll m				
George McCrae -- "Rock Your Baby" ~ Sexy mama, Woman, take me in your arms, Rock your baby, Woman, take me in your arms, Rock your ba				
Roger Whittaker -- "The Last Farewell" ~ There's a ship lies rigged and ready in the harbor, Tomorrow for old England she sails, Far.				
ABBA -- "Fernando" ~ Can you hear the drums Fernando, I remember long ago another starry night like this, In the firelight Fernando,				
John Travolta & Olivia Newton-John -- "You're the One That I Want" ~ I got chills, They're multiplying, And I'm losing control, Cause				
Gloria Gaynor -- "I Will Survive" ~ At first I was afraid, I was petrified, Kept thinking I could never live without you by my side, .				
Village People -- "Y.M.C.A." ~ Young man, there's no need to feel down, I said, young man, pick yourself off the ground, I said, youn				
Trio -- "Da Da Da" ~ Aha aha aha, Aha, aha, aha, Was ist los mit dir mein Schatz? Aha, Geht es immer nur bergab? Aha, Geht nur das wa				
Band Aid -- "Do They Know It's Christmas?" ~ It's Christmas time, there's no need to be afraid, At Christmas time, we let in light an				
USA for Africa -- "We Are the World" ~ There comes a time, When we heed a certain call, When the world must come together as one, The				
Kaoma -- "Lambada" ~ Chorando se foi quem um dia se fez chorar, Chorando se foi quem um dia se fez chorar, Chorando estara, ad				
Bryan Adams -- "(Everything I Do) I Do It for You" ~ Look into my eyes, You will see, what you mean to me, Search your heart, search				
Scorpions -- "Wind of Change" ~ I follow the Moskva, Down to Gorky Park, Listening to the wind of change, An August summer night, Sol				
Whitney Houston -- "I Will Always Love You" ~ If I should stay, I'll only be in your way, So I'll go, but I know I'll, Think of you e				
Los del Rio -- "Macarena" ~ Dale a tu cuerpo alegria Macarena, Que tu cuerpo es pa' darle alegria y cosa buena, Dale a tu cuerpo aleg				
Andrea Bocelli & Sarah Brightman -- "Time to Say Goodbye" ~ Quando sono sola, Sogno all'orizzonte, E mancan le parole, Si' lo so che.				
Toni Braxton -- "Un-Break My Heart" ~ Don't leave me in all this pain, Don't leave me out in the rain, Come back and bring back my.				
Elton John -- "Candle in the Wind 1997"/"Something About the Way You Look Tonight" ~ Goodbye Norma Jean, Though I never knew you at a				
Celine Dion -- "My Heart Will Go On" ~ Every night in my dreams, I see you, I feel you, That is how I know you, go on, Far across the				
Cher -- "Believe" ~ No matter how hard I try, You keep pushing me aside, And I can't break through, There's no talking to you, It's s				
Panjabi MC -- "Mundian To Bach Ke" ~ It's the Roc in the building, Calib, Ramei, Tarrell in the house, Minian to but the tiri bachin				
Britney Spears -- "...Baby One More Time" ~ Oh baby, baby, how was I supposed to know, That something wasn't right here, Oh baby, bab				
Celine Dion & Andrea Bocelli -- "The Prayer" ~ I pray you'll be our eyes and watch us where we go, And help us to be wise in times wh				

Best-Selling Singles more than 10 million copies, sorted by release date

Bar Chart of Text

No: "Christmas" beats love (but occurs in only 2 songs)

Bing Crosby -- "Silent Night" ~ Silent night, Holy night, All is calm, All is bright, Round your virgin mother and child, Holy infant, the Ink Spots -- "If I Didn't Care" ~ If I didn't care more than words can say, If I didn't care, would I feel this way? If this isn't				
Bing Crosby -- "White Christmas" ~ I'm dreaming of a white Christmas, Just like the ones we used to know, Where the treetops glisten				
Roy Acuff -- "Wabash Cannonball" ~ From the great Atlantic ocean to the wide Pacific shore, From the queen of flowing mountain to the				
Wills Brothers -- "Paper Doll" ~ I'm gonna buy a paper doll that I can call my own, A doll that other fellows cannot steal, And their				
Gene Autry -- "Rudolph the Red-Nosed Reindeer" ~ You know Dasher and Dancer and Prancer and Vixen, Comet and Cupid and Donner and Blitzen				
Patti Page -- "Tennessee Waltz" ~ I was dancing with my darling to the Tennessee Waltz, When an old friend I happened to see, I introduced				
Bill Haley & His Comets -- "Rock Around the Clock" ~ One, two, three o'clock, four o'clock, rock, Five, six, seven o'clock, eight o'clock				
The Penguins -- "Earth Angel" ~ Earth angel, earth angel, Will you be mine?, My darling dear, Love you all the time, I'm just a fool,				
Elvis Presley -- "Hound Dog" ~ Hey baby I'm here to tell you about yourself, You ain't nothin' but a hound dog, scratchin' all the time				
Paul Anka -- "Diana" ~ I'm so young and you're so old, This, my darling, I've been told, I don't care just what they say, 'Cause fore and				
Domenico Modugno -- "Nel blu dipinto di blu (Volare)" ~ Sometimes the world is a valley of heartaches and tears, And in the hustle and				
Elvis Presley -- "It's Now or Never" ~ It's now or never, Come hold me tight, Kiss me my darling, Be mine tonight, Tomorrow will be				
Kyu Sakamoto -- "Sukiyaki" ~ Let's look up as we walk, so that the tears don't spill, Remembering that spring day, that lonely night.				
The Beatles -- "I Want to Hold Your Hand" ~ Oh yeah I tell you somethin', I think you'll understand, When I say that somethin', I want				
The Monkees -- "I'm a Believer" ~ I thought love was only true in fairy tales, Meant for someone else but not for me, Love was out to				
Procol Harum -- "A Whiter Shade of Pale" ~ We skipped the light fandango, Turned cartwheels 'cross the floor, I was feeling kinda sea				
Mungo Jerry -- "In the Summertime" ~ In the summertime when the weather is hot, You can stretch right up and touch the sky, When the				
Middle of the Road -- "Chirpy Chirpy Cheep Cheep" ~ Where's your mamma gone, Little baby bird, Where's your mamma gone, Far far away.				
Carl Douglas -- "Kung Fu Fighting" ~ Now here it is, one to make you move, Something with a funky kung fu groove, Something that'll make				
George McCrae -- "Rock Your Baby" ~ Sexy mama, Woman, take me in your arms, Rock your baby, Woman, take me in your arms, Rock your baby				
Roger Whittaker -- "The Last Farewell" ~ There's a ship lies rigged and ready in the harbor, Tomorrow for old England she sails, Far				
ABBA -- "Fernando" ~ Can you hear the drums Fernando, I remember long ago another starry night like this, In the firelight Fernando,				
John Travolta & Olivia Newton-John -- "You're the One That I Want" ~ I got chills, They're multiplying, And I'm losing control, Cause				
Gloria Gaynor -- "I Will Survive" ~ At first I was afraid, I was petrified, Kept thinking I could never live without you by my side,				
Village People -- "Y.M.C.A." ~ Young man, there's no need to feel down, I said, young man, pick yourself off the ground, I said, young				
Trio -- "Da Da Da" ~ Aha aha aha, Aha, aha, aha, Was ist los mit dir mein Schatz? Aha, Geht es immer nur bergab? Aha, Geht nur das was				
Band Aid -- "Do They Know It's Christmas?" ~ It's Christmas time, there's no need to be afraid, At Christmas time, we let in light an				
USA for Africa -- "We Are the World" ~ There comes a time, When we heed a certain call, When the world must come together as one, The				
Kaoma -- "Lambada" ~ Chorando se foi quem um dia so me fez chorar, Chorando se foi quem um dia so me fez chorar, Chorando estara, ao				
Bryan Adams -- "(Everything I Do) I Do It for You" ~ Look into my eyes, You will see, what you mean to me, Search your heart, search				
Scorpions -- "Wind of Change" ~ I follow the Moskva, Down to Gorky Park, Listening to the wind of change, An August summer night, Sol				
Whitney Houston -- "I Will Always Love You" ~ If I should stay, I'll only be in your way, So I'll go, but I know I'll, Think of you every				
Los del Rio -- "Macarena" ~ Dale a tu cuerpo alegria Macarena, Que tu cuerpo es pa' darle alegria y cosa buena, Dale a tu cuerpo aleg				
Andrea Bocelli & Sarah Brightman -- "Time to Say Goodbye" ~ Quando sono sola, Sogno all'orizzonte, E mancan le parole, Si' lo so che.				
Toni Braxton -- "Un-Break My Heart" ~ Don't leave me in all this pain, Don't leave me out in the rain, Come back and bring back my				
Elton John -- "Candle in the Wind 1997"/"Something About the Way You Look Tonight" ~ Goodbye Norma Jean, Though I never knew you at				
Celine Dion -- "My Heart Will Go On" ~ Every night in my dreams, I see you, I feel you, That is how I know you, go on, Far across the				
Cher -- "Believe" ~ No matter how hard I try, You keep pushing me aside, And I can't break through, There's no talking to you, It's so				
Panjabi MC -- "Mundian To Bach Ke" ~ It's the Roc in the building, Calib, Ramei, Tarrell in the house, Minian to but the tiri bachin				
Britney Spears -- "...Baby One More Time" ~ Oh baby, baby, how was I supposed to know, That something wasn't right here, Oh baby, baby				
Celine Dion & Andrea Bocelli -- "The Prayer" ~ I pray you'll be our eyes and watch us where we go, And help us to be wise in times wh				

Bar Chart of Text

And "baby" beats both in counts and songs

Best-Selling Singles more than 10 million copies, sorted by release date

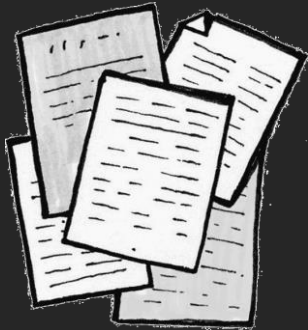
Bing Crosby -- "Silent Night" ~ Silent night, Holy night, All is calm, All is bright, Round your virgin mother and child, Holy infant the Inks Spots -- "If I Didn't Care" ~ If I didn't care more than words can say, If I didn't care, would I feel this way? If this isn't	10	20	30	40	50m
Bing Crosby -- "White Christmas" ~ I'm dreaming of a white Christmas, Just like the ones we used to know, Where the treetops glisten					
Roy Acuff -- "Wabash Cannonball" ~ From the great Atlantic ocean to the wide Pacific shore, From the queen of flowing mountain to the					
Wills Brothers -- "Paper Doll" ~ I'm gonna buy a paper doll that I can call my own, A doll that other fellows cannot steal, And then					
Gene Autry -- "Rudolph the Red-Nosed Reindeer" ~ You know Dasher and Dancer and Prancer and Vixen, Comet and Cupid and Donner and Bl					
Patti Page -- "Tennessee Waltz" ~ I was dancing with my darling to the Tennessee Waltz, When an old friend I happened to see, I intro					
Bill Haley & His Comets -- "Rock Around the Clock" ~ One, two, three o'clock, four o'clock, rock, Five, six, seven o'clock, eight o'clock					
The Penguins -- "Earth Angel" ~ Earth angel, earth angel, Will you be mine?, My darling dear, Love you all the time, I'm just a fool,					
Elvis Presley -- "Hound Dog" ~ Hey baby I'm here to tell you about yourself, You ain't nothin' but a hound dog, scratchin' all the ti					
Paul Anka -- "Diana" ~ I'm so young and you're so old, This, my darling, I've been told, I don't care just what they say, 'Cause fore					
Domenico Modugno -- "Nel blu dipinto di blu (Volare)" ~ Sometimes the world is a valley of heartaches and tears, And in the hustle an					
Elvis Presley -- "It's Now or Never" ~ It's now or never, Come hold me tight, Kiss me my darling, Be mine tonight, Tomorrow will be t					
Kyu Sakamoto -- "Sukiyaki" ~ Let's look up as we walk, so that the tears don't spill, Remembering that spring day, that lonely night.					
The Beatles -- "I Want to Hold Your Hand" ~ Oh yeah I tell you somethin', I think you'll understand, When I say that somethin', I wan					
The Monkees -- "I'm a Believer" ~ I thought love was only true in fairy tales, Meant for someone else but not for me, Love was out to					
Procol Harum -- "A Whiter Shade of Pale" ~ We skipped the light fandango, Turned cartwheels 'cross the floor, I was feeling kinda sea					
Mungo Jerry -- "In the Summertime" ~ In the summertime when the weather is hot, You can stretch right up and touch the sky, When the					
Middle of the Road -- "Chirpy Chirpy Cheep Cheep" ~ Where's your mamma gone, Little baby bird, Where's your mamma gone, Far far away.					
Carl Douglas -- "Kung Fu Fighting" ~ Now here it is, one to make you move, Something with a funky kung fu groove, Something that'll m					
George McCrae -- "Rock Your Baby" ~ Sexy mama, Woman, take me in your arms, Rock your baby, Woman, take me in your arms, Rock your ba					
Roger Whittaker -- "The Last Farewell" ~ There's a ship lies rigged and ready in the harbor, Tomorrow for old England she sails, Far.					
ABBA -- "Fernando" ~ Can you hear the drums Fernando, I remember long ago another starry night like this, In the firelight Fernando,					
John Travolta & Olivia Newton-John -- "You're the One That I Want" ~ I got chills, They're multiplying, And I'm losing control, Cause					
Gloria Gaynor -- "I Will Survive" ~ At first I was afraid, I was petrified, Kept thinking I could never live without you by my side, I					
Village People -- "Y.M.C.A." ~ Young man, there's no need to feel down, I said, young man, pick yourself off the ground, I said, youn					
Trio -- "Da Da Da" ~ Aha aha aha, Aha, aha, aha, Was ist los mit dir mein Schatz? Aha, Geht es immer nur bergab? Aha, Geht nur das wa					
Band Aid -- "Do They Know It's Christmas?" ~ It's Christmas time, there's no need to be afraid, At Christmas time, we let in light an					
USA for Africa -- "We Are the World" ~ There comes a time, When we heed a certain call, When the world must come together as one, The					
Kaoma -- "Lambada" ~ Chorando se foi quem um dia se me fez chorar, Chorando se foi quem um dia se me fez chorar, Chorando estara, ad					
Bryan Adams -- "(Everything I Do) I Do It for You" ~ Look into my eyes, You will see, what you mean to me, Search your heart, search					
Scorpions -- "Wind of Change" ~ I follow the Moskva, Down to Gorky Park, Listening to the wind of change, An August summer night, Sol					
Whitney Houston -- "I Will Always Love You" ~ If I should stay, I'll only be in your way, So I'll go, but I know I'll, Think of you e					
Los del Rio -- "Macarena" ~ Dale a tu cuerpo alegria Macarena, Que tu cuerpo es pa' darle alegria y cosa buena, Dale a tu cuerpo aleg					
Andrea Bocelli & Sarah Brightman -- "Time to Say Goodbye" ~ Quando sono sola, Sogno all'orizzonte, E mancan le parole, Si' lo so che.					
Toni Braxton -- "Un-Break My Heart" ~ Don't leave me in all this pain, Don't leave me out in the rain, Come back and bring back my.					
Elton John -- "Candle in the Wind 1997"/"Something About the Way You Look Tonight" ~ Goodbye Norma Jean, Though I never knew you at a					
Celine Dion -- "My Heart Will Go On" ~ Every night in my dreams, I see you, I feel you, That is how I know you, go on, Far across the					
Cher -- "Believe" ~ No matter how hard I try, You keep pushing me aside, And I can't break through, There's no talking to you, It's s					
Panjabi MC -- "Mundian To Bach Ke" ~ It's the Roc in the building, Calib, Ramei, Tarrell in the house, Minian to but the tiri bachin					
Britney Spears -- "...Baby One More Time" ~ Oh baby, baby, how was I supposed to know, That something wasn't right here, Oh baby, bab					
Celine Dion & Andrea Bocelli -- "The Prayer" ~ I pray you'll be our eyes and watch us where we go, And help us to be wise in times wh					

3. *But what about*
BIG DATA TEXT ANALYTICS?

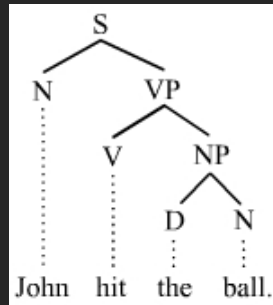
Text Analytics 101

Typical “Natural Language Processing” Pipeline

Raw
Text



Preprocessing



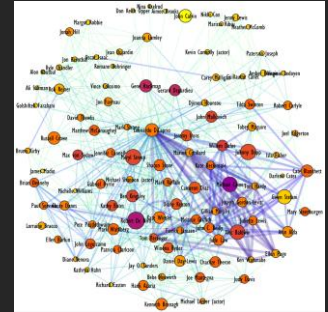
https://en.wikipedia.org/wiki/Parse_tree

Numerical
Representation

	Document 1	Document 2	Document 3	Document 4	Document 5	Document 6	Document 7	
Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	0	3	0	0

<http://brandonrose.org/clustering>




Analysis



Brath and Jonker. Graph Analysis and Visualization. 2015. Wiley.
<http://ca.wiley.com/WileyCDA/WileyTitle/productCd-1118845846.html>

Preprocessing Raw Text

Tokenize  words "Hello, I'm Dr. Jones."
sentences ['Hello', 'I', 'm' 'Dr' 'Jones']

Normalize  stem ponies poni
 lemmatize ponies pony
 lowercase Ponies ponies

Remove Noise  stop words John's car is red, right?
punctuation **John's car is red, right?**

Document Representation

Counting

word count
sentence count
syllable count

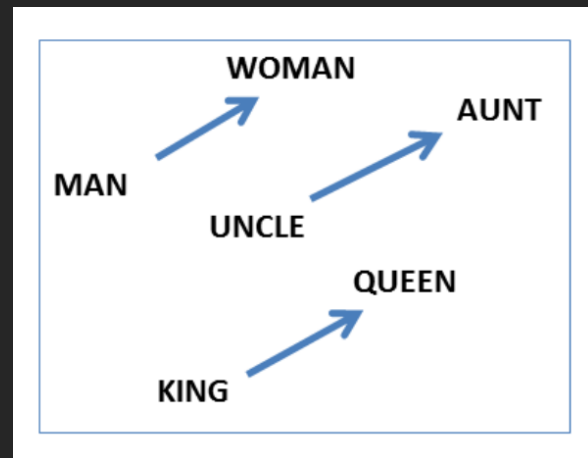
TFIDF

times
term t is
in doc d $\times \frac{1}{\# \text{ docs term } t \text{ appears in}}$

Word Embedding

Text Vectors

	king	queen	man	woman
count	32	147	14	83
TFIDF	0.001	0.097	0.051	0.147



Mikolov et al., NAACL HLT, 2013. <https://www.aclweb.org/anthology/N13-1090>

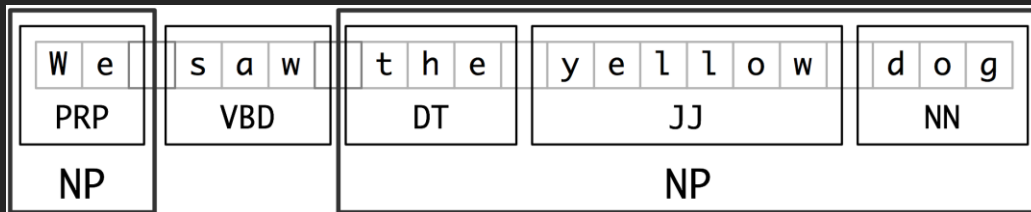
Document Representation

Part-of-Speech tagging

['And, 'now', 'for', 'something', 'completely', 'different']

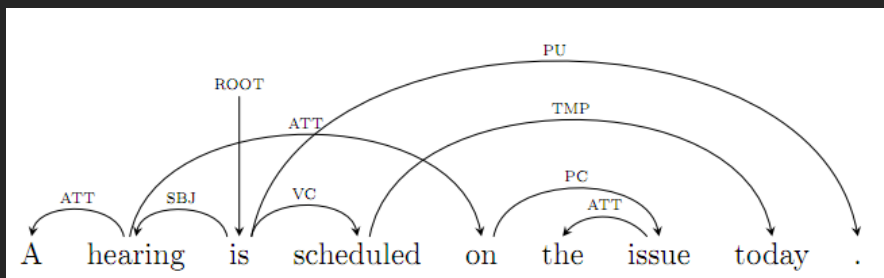
'CC' 'RB' 'IN' 'NN' 'RB' 'JJ'

Chunking / Phrase extraction



Bird, Steven, Edward Loper and Ewan Klein (2009). Natural Language Processing with Python. O'Reilly <http://www.nltk.org/book/ch07.html>

Dependency parse tree



<https://en.wikibooks.org/wiki/LaTeX/Linguistics#/media/File:Latex-dependency-parse-example-with-tikz-dependency.png>

Text Analysis

Topic Modeling

Using statistical features to discover abstract 'topics' that occur in a collection of documents

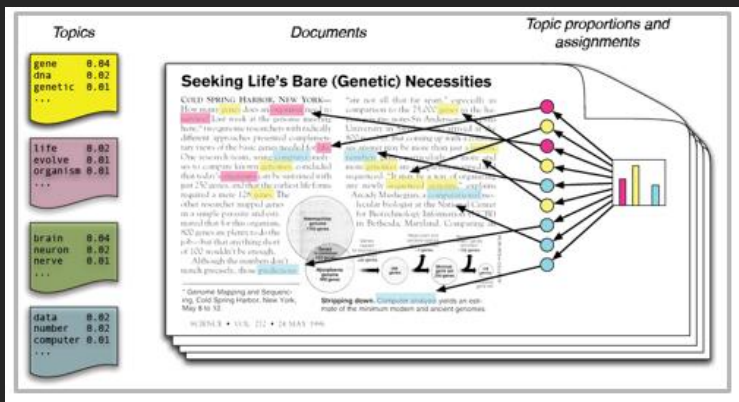
'topic' is a probability distribution of words

Topic 1: algorithm, computer, data, ...

Topic 2: gene, data, bio, ...

Topic 3: math, equation, matrix, ...

Topic 4: chemical, carbon, water, ...



Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77-84. <http://www.cs.columbia.edu/~blei/papers/Blei2012.pdf> or <https://lists.cs.princeton.edu/pipermail/topic-models/2011-August/001535.html>



Topic graph, showing topic nodes and edge connections removed. Brath and Jonker. Graph Analysis and Visualization. 2015. Wiley. <http://ca.wiley.com/WileyCDA/WileyTitle/productCd-1118846846.html>

Other NLP Tasks



Coreference resolution	"Scott took a flight. He said it was very crowded"
Text Classification	Spam / Not Spam
Text Similarity	(definitely, definately) = 1
Sentiment Analysis	Strata was great this year! :) = POSITIVE
Entity extraction	New York is a metropolis → ('New York', 'city')
Text Summarization	Mostly harmless
Machine Translation	Mostly harmless -> ほとんど無害
Natural Language Generation	"Sounds good, talk to you then"

4. TEXT ANALYTICS & VIZ

together!

NLP to find Traits

1. Extract the characters
2. Extract the adjectives near the characters
3. Visualize

e.g. This example, one sentence from *Grimms' Fairy Tales*.

When you come to the old king, and he asks for the beautiful princess, you must say, "Here she is!"



When you come to the old **king**, and he asks for the beautiful **princess**, you must say, "Here she is!"



When you come to the old **king**, and he asks for the beautiful **princess**, you must say, "Here she is!"

Traits

Apply to
whole text

Unexpected
character
traits?

Are witches really
wicked, or maybe
just old?

Characters from Grimms' Fairy Tales with Associated Adjectives Weighted by Frequency

Character List of adjectives, weighted by frequency: 2 3 4-5 6-9 10+

bird	beautiful	splendid	open	wooden	like	hanging				
cat	little	one	long							
fox	old	dead	young	first	fast					
gretel	little	poor	good							
hansel	little	like	fat							
bride	false	true	first	right	real					
king	old	great	one	young	three	angry	beautiful	married	like	third
princess	beautiful	young	last	dear	enchanted	strange	true	third	free	
girl	little	poor	lazy	pretty	young	dead	beautiful	silly		
queen	beautiful	late	little	far						
wife	standing	married	next	poor	two	beautiful	new	one	dear	true
witch	old	wicked								

0

5

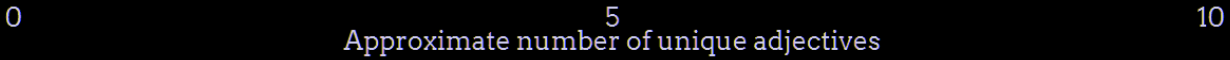
10

Approximate number of unique adjectives

CEO's Use of Emotion Words During Earnings Calls, weighted by frequency

CEO List of adjectives, weighted by frequency: 2 3-4 5-7 8-11 12+

Carl Bass	pleased good
Carl Trowell	demand lower good debt improve cash older grow tender retirement late termination
Carlos A. Rodriguez	grow confidence pleased success excited proud progress
Carol M. Meyrowitz	pleased confident share margin grow powerful organization excellent shopping terrific passionate gain
Charles Alutto	good kind waste cash excited lines share safe hazardous versus fee pleased resources tough respo
Charles B. Stanley	completion improvement lower gross progress reject rejection
Charles Bunch	improvement lower share cash versus margin unfavorable gross good aggressive tax income kind i
Charlie Scharf	kind good share change lower grow pay pleased cross case versus helpful terrific progress excited
Chris Connor	share kind margin good pretty demand gross john confidence versus income cash bad interior ren
Chris Sinclair	good progress pretty hot content grow share change organization improvement gross finally kind e
Christopher Crane	good create late hearing grow successful pleased share advocacy regulatory clean majority kind debt
Chuck Bunch	good improvement share unfavorable versus cash lower excellent demand respect difficult debt inc
Chuck Jones	case cash good kind improvement progress finally cutting inflation court improve share risk electric
Clay Williams	lower demand pretty kind good excluding improve completion versus share grow deflation tough di
Clifton Pemble	margin lower income gross finally excited share pretty
Craig A. Menear	share pleased



CEO's Use of Emotion Words During Earnings Calls, weighted by frequency

CEO List of adjectives, weighted by frequency: 2 3-4 5-7 8-11 12+

Andrew P. Wilson	battlefield star console entertainment dragon fun ultimate
Bruce Broussard	star tax enforce good progress successful
David Hallal	disease complement devastating progress therapeutic success majority hospital
Fabrizio Freda	beauty exciting prestige successful
Gracia Martore	love content respect terrific apologize lagging
J.P. Bilbrey	chocolate shanghai lower unfavorable force
Jim Tisch	diamond cash income special pay doubt lose
John D. Idol	luxury demand shopping watch bottom
John Ferriola	pig improvement cash lower challenge
John Lechleiter	cancer disease deal regulatory tax therapeutic
Kelly King	mortgage risk income tough money fee
Larry D. Young	punch case music football love passion
Les Moonves	content deal bang success terrific confident scorpion late
Mark Parker	basketball demand excited elite offense discipline
Mark R. Hunter	beer tax lower unfavorable loss organization banquet blue mad
Omid Kordestani	shopping holiday progress pay music watch friction
Richard J. Harshman	demand fell defect excellent labor wages aggressive falling
Steven Temares	sue shopping baby improve excited exciting share success pleased fulfillment
Thomas Kennedy	missile defense demand government award bottom threat

Electronic Arts

Estée Lauder

Alexion Pharmaceuticals

Loews

0

5

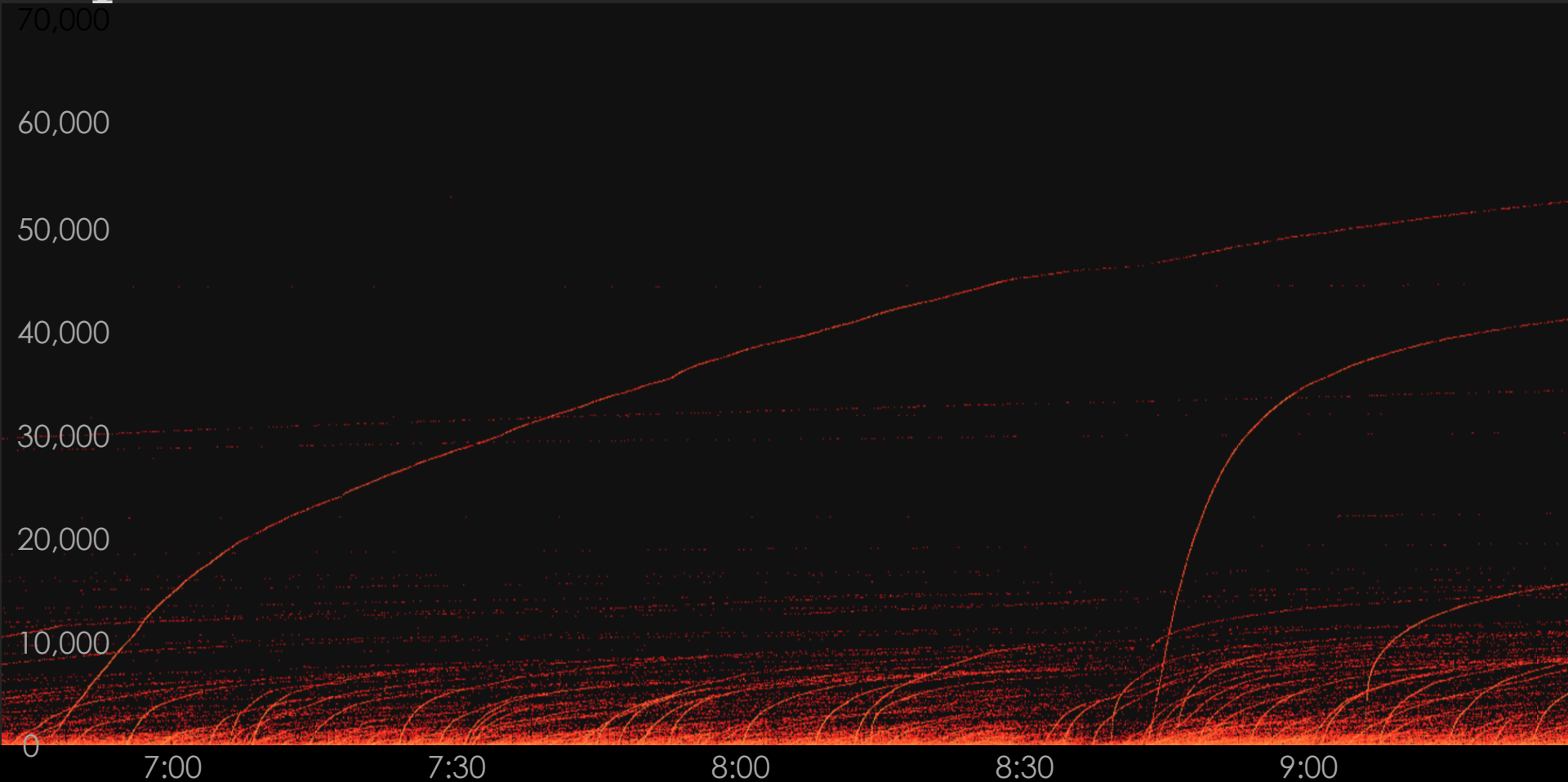
10

5. **STICKY** social media

Plot a billion tweets



Superbowl 50 Retweets



Uncharted Flare Chart. Vertical axis is number of retweets. Lines are individual tweets. Some tweets flare up quickly then flatten, some go up, stutter, then go up more, etc..

Superbowl 50 Retweets

uncharted



Note: some retweet values are **offscreen**



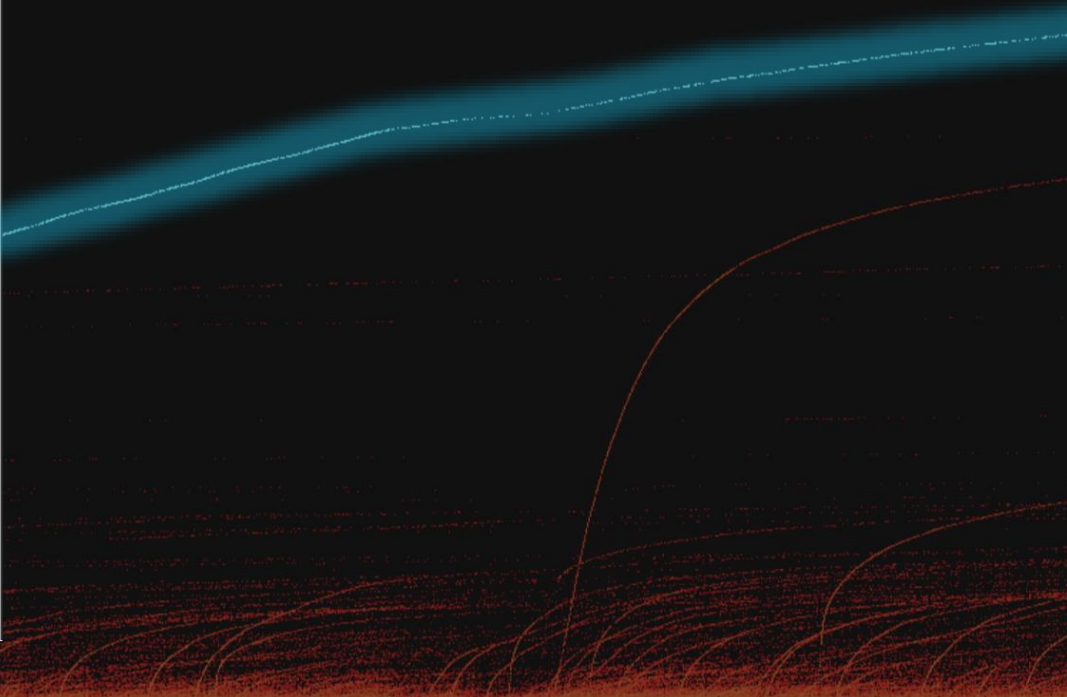
 **Betty White** 
@BettyMWhite

 Follow

I taught @CameronNewton everything he knows. #SuperBowl

7:44 PM - Feb 7, 2016

 919  75,417  87,997



7:00

7:30


8:00


8:30

9:00

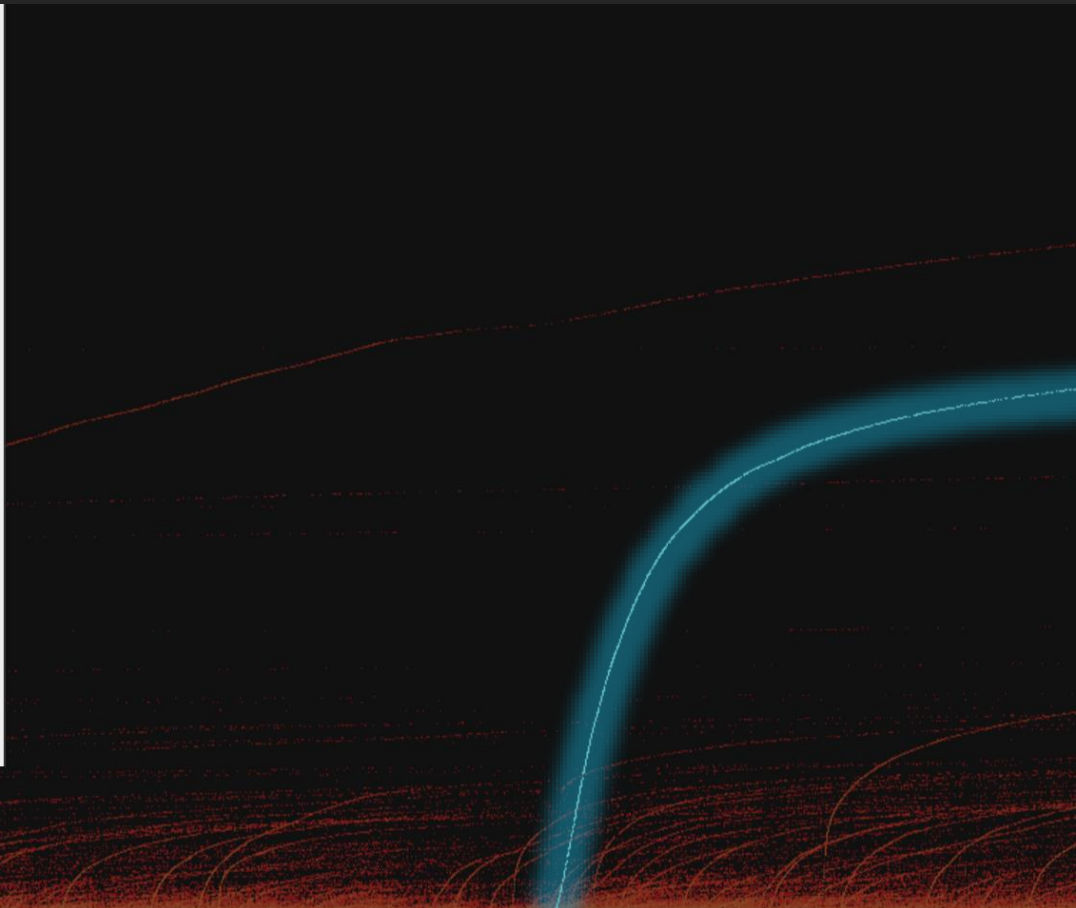
Superbowl 50 Retweets

uncharted



Coldplay 
@coldplay [Follow](#)

#SB50 #PepsiHalftime
9:42 PM - Feb 7, 2016
751 49,474 58,172



7:00

7:30

8:00

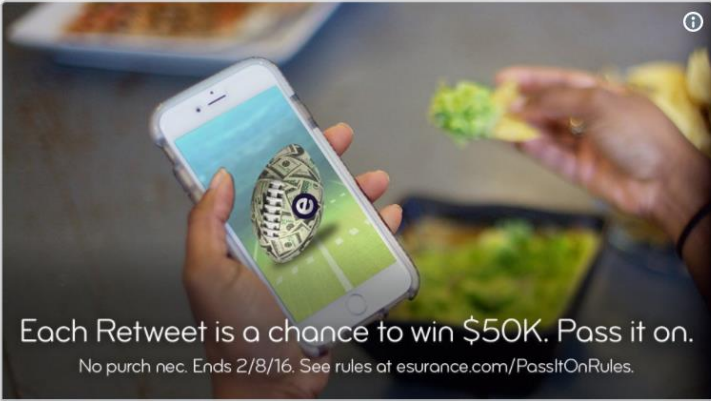
8:30

9:00




Uncharted Flare Chart. Coldplay has a popular tweet at 8:45 pm which then flattens out.

Superbowl 50 Retweets

uncharted






Each Retweet is a chance to win \$50K. Pass it on.
No purch nec. Ends 2/8/16. See rules at esurance.com/PassItOnRules.

 **Esurance** 
@esurance 

Before you reach for another chip, RT for your chance to win.
[#EsuranceSweepstakes](#) [#SB50](#)

9:12 PM - Feb 7, 2016

 476  12,500  1,924

7:00

7:30

8:00

8:30

9:00

Uncharted Flare Chart. Esurance has a retweet campaign which pops up quickly but starts to flatten after 20 minutes.

Superbowl 50 Retweets

uncharted ×

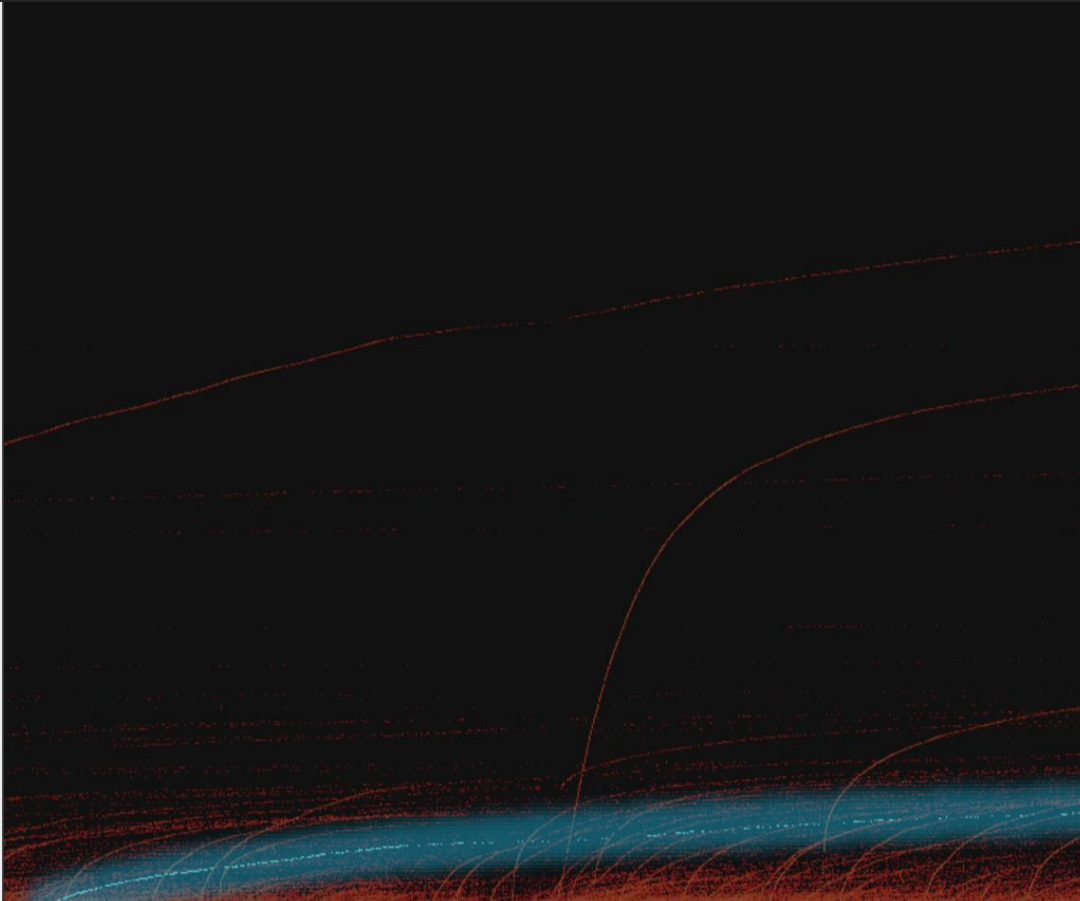


 **Scott Kelly** 
@StationCDRKelly 

Got to see the #SuperBowl in person after all! But at 17,500MPH, it didn't last long. #YearInSpace

8:58 PM - Feb 7, 2016

 445  11,831  22,973



7:00

7:30

8:00

8:30

9:00

Uncharted Flare Chart. This tweet from the International Space Station doesn't flare up, but keeps getting more and more retweets. Note how only one tweet can be examined at a time.

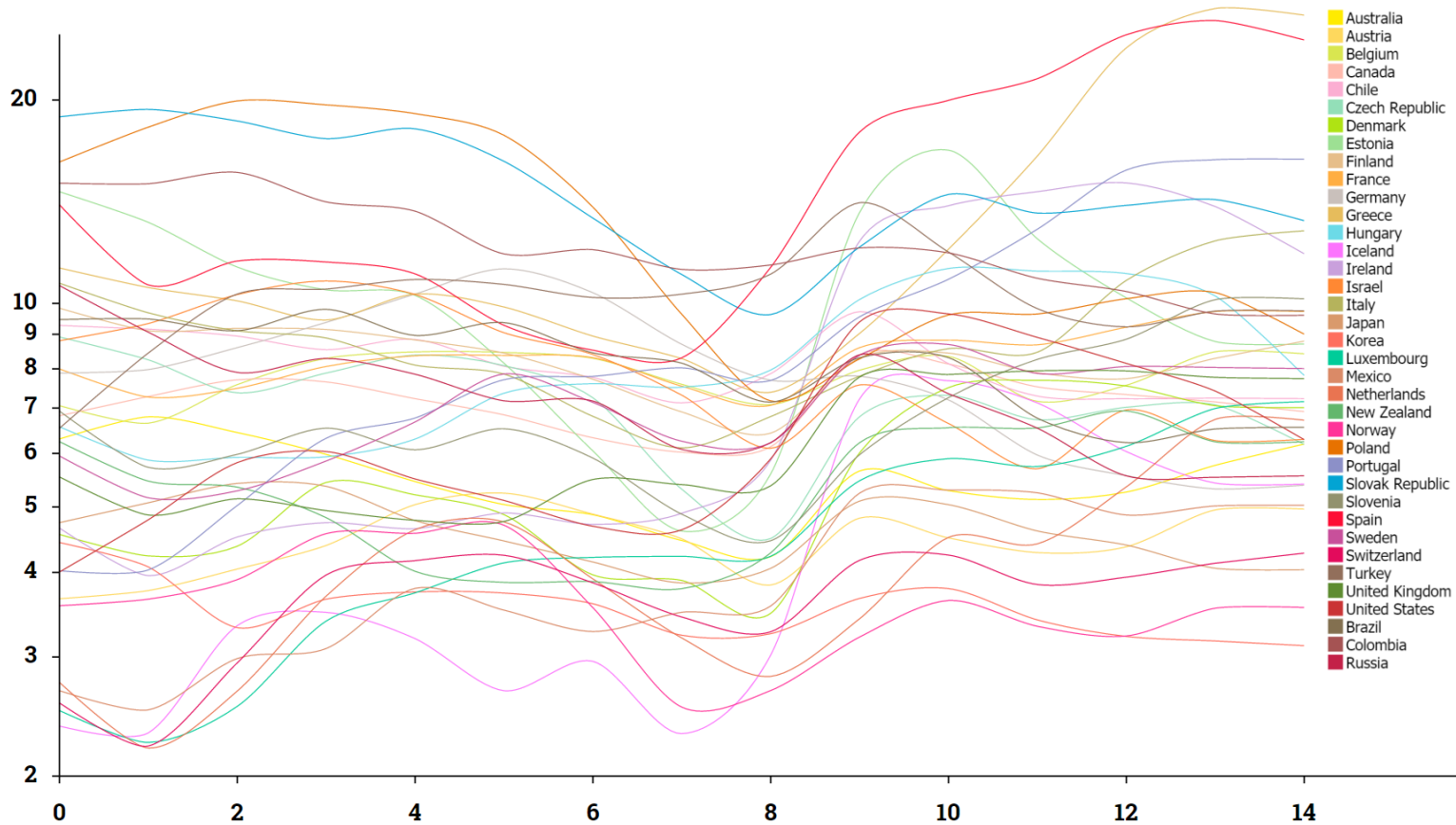
Line Chart

- Which line is which?

Similar problem: many lines on a line chart are difficult to disambiguate. Only way to disambiguate is to interact with each line individually. Slow!

Unemployment Rate as Percent Civilian Labour Force

2000-2014 by Country.



Source: OECD (stats.oecd.org, retrieved Oct 24, 2015). Author: Richard Brath/ London South Bank University 2015 (richardbrath.wordpress.com)

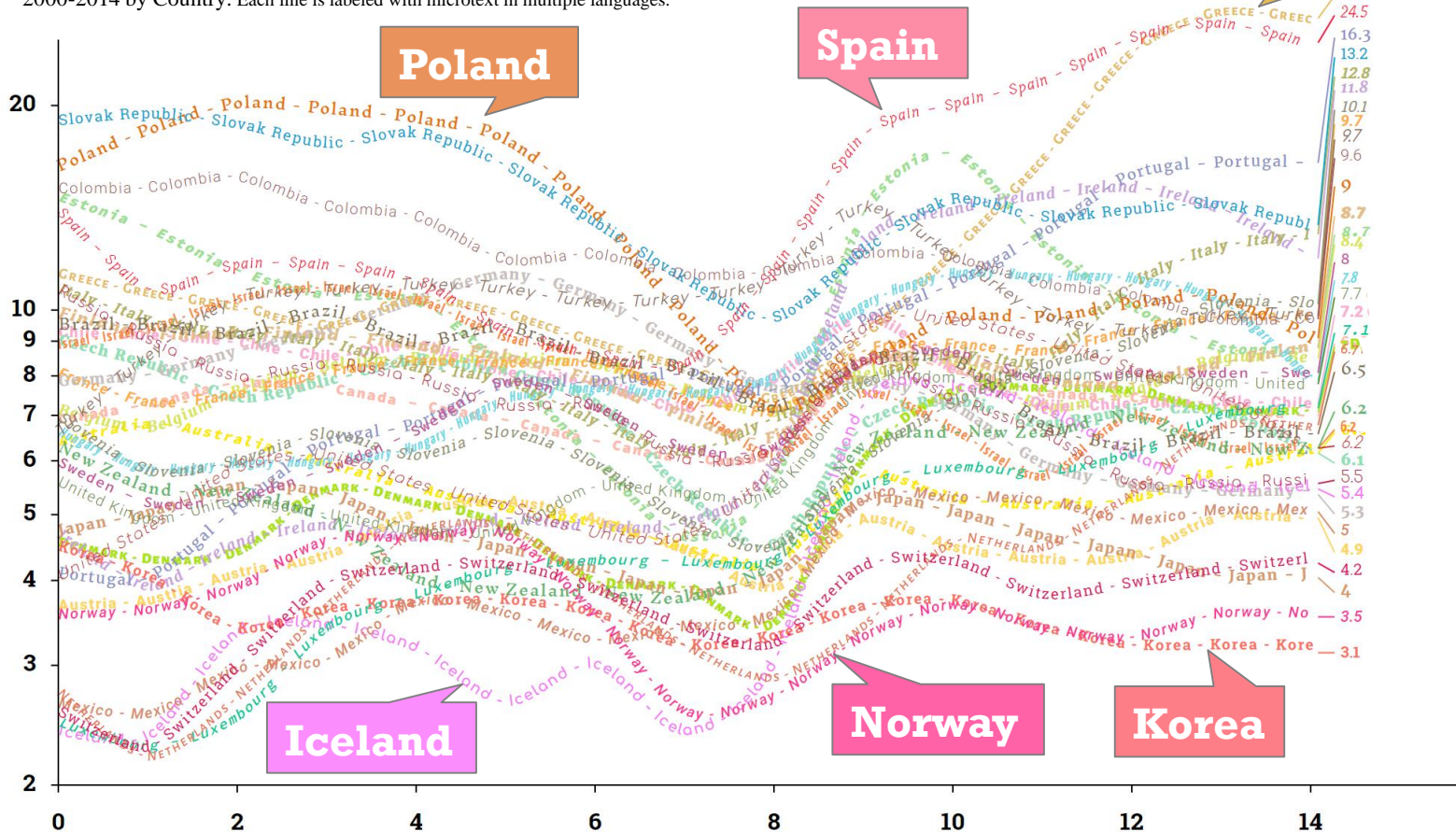
Unemployment Rate as Percent Civilian Labour Force

Greece

2000-2014 by Country. Each line is labeled with microtext in multiple languages.

Line Chart

Turn the line into text



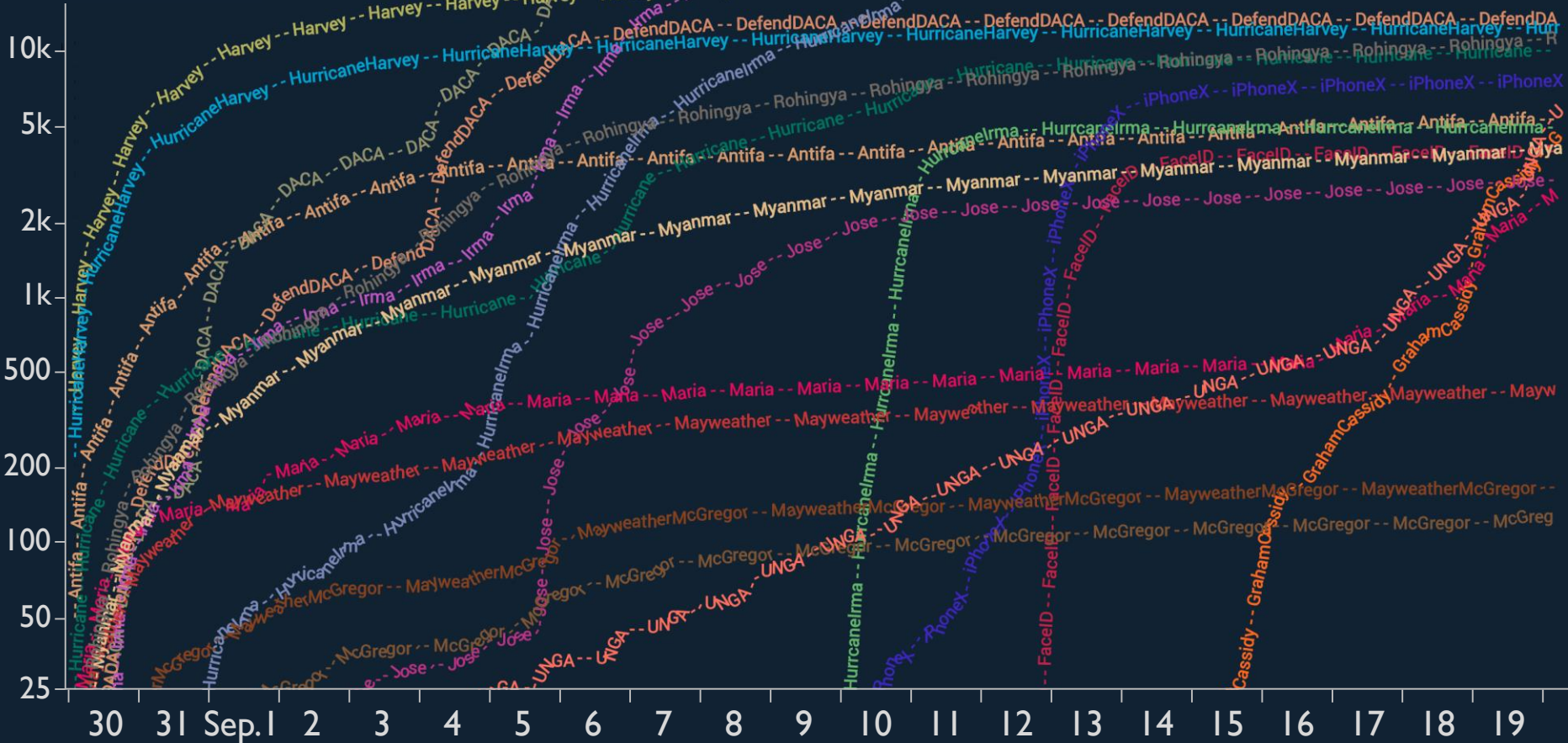
Source: OECD (stats.oecd.org, retrieved Oct 24, 2015). Author: Richard Brath/ London South Bank University 2015 (richardbrath.wordpress.com)

When lines are text, identity can be directly read. Callouts for audience at back of room.

Top Hashtags

Sept 2017

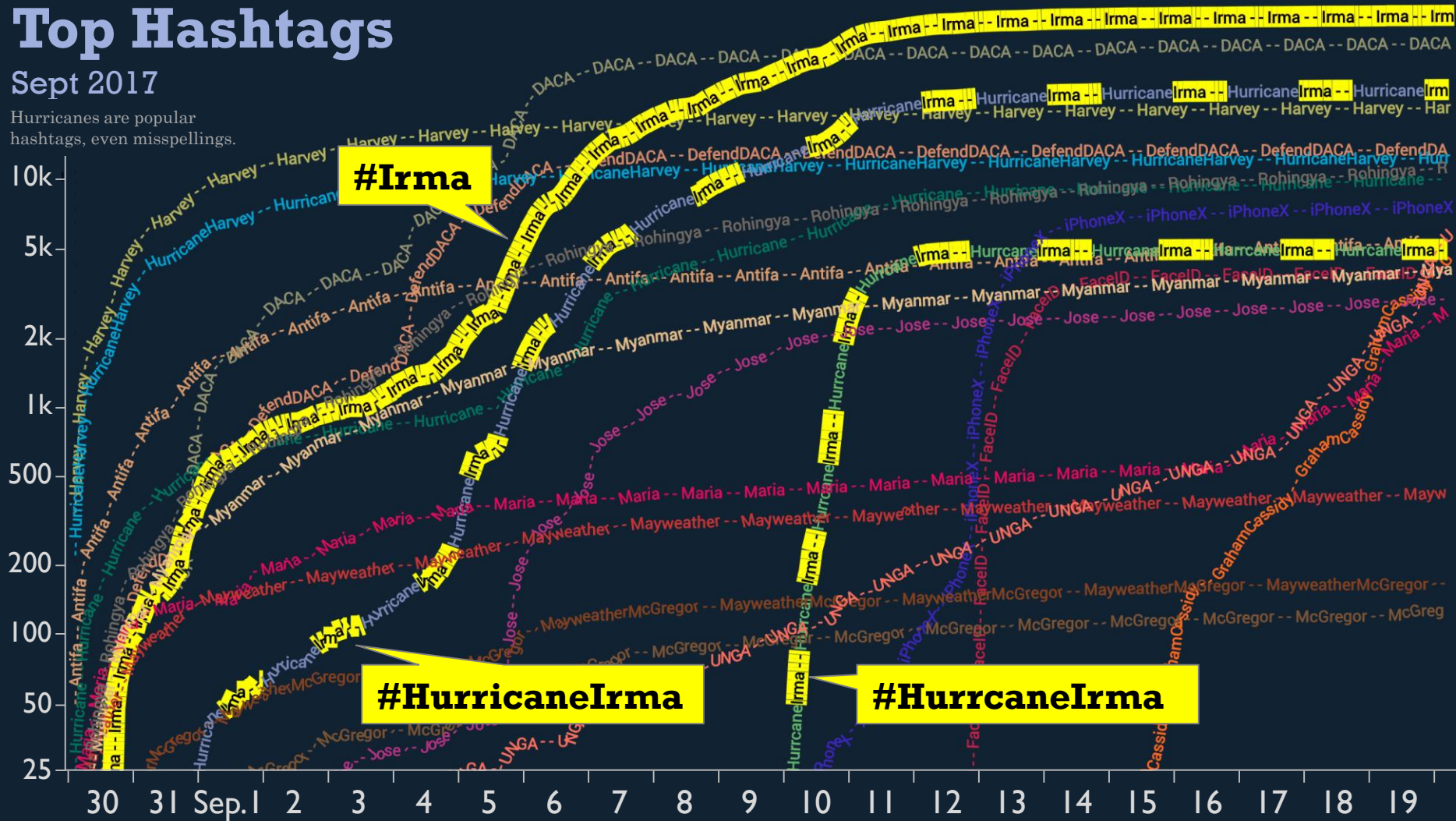
Flare Chart of Hashtags



Top Hashtags

Sept 2017

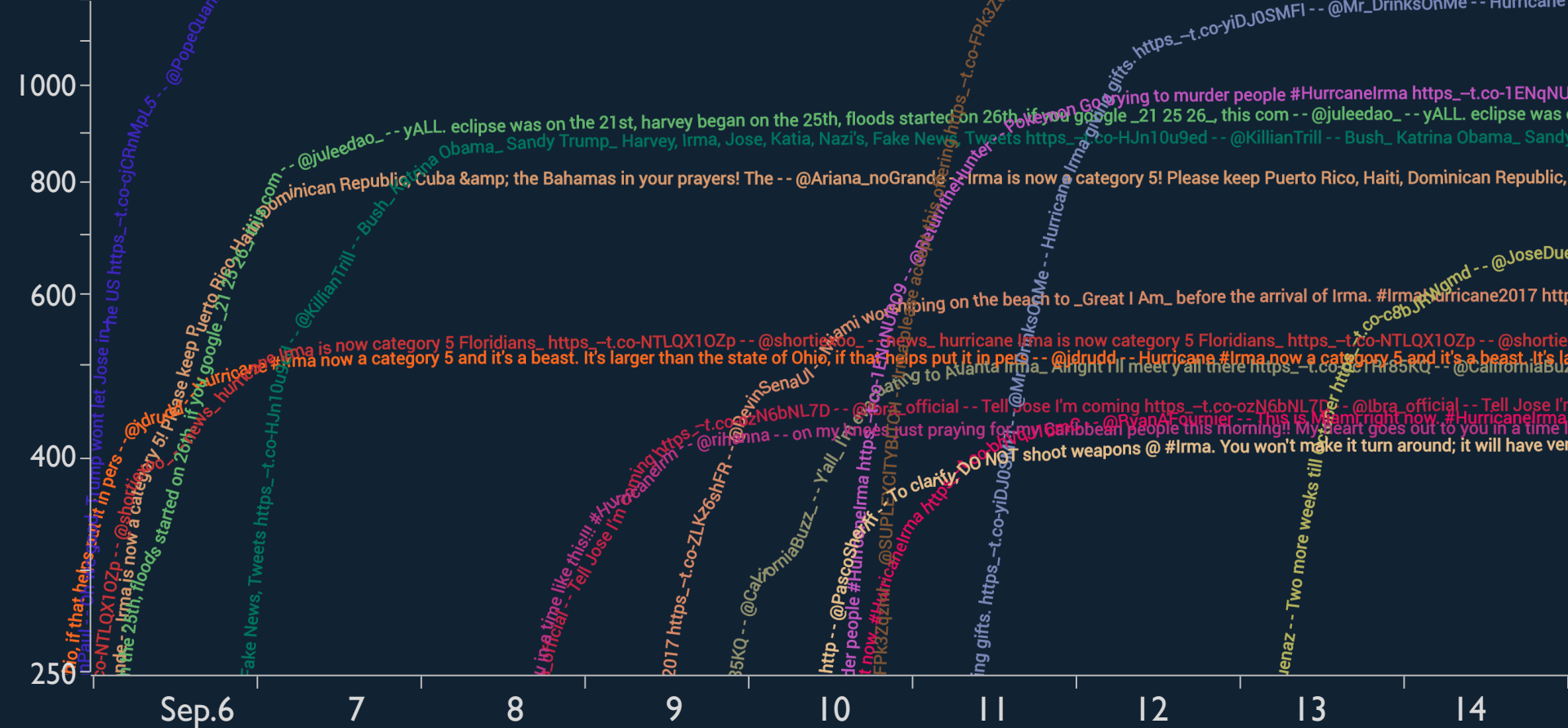
Hurricanes are popular hashtags, even misspellings.



Top Hurricane Retweets

Sep. 6-14, 2017

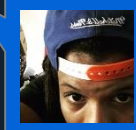
Detailed hurricane tweet content on flare chart.



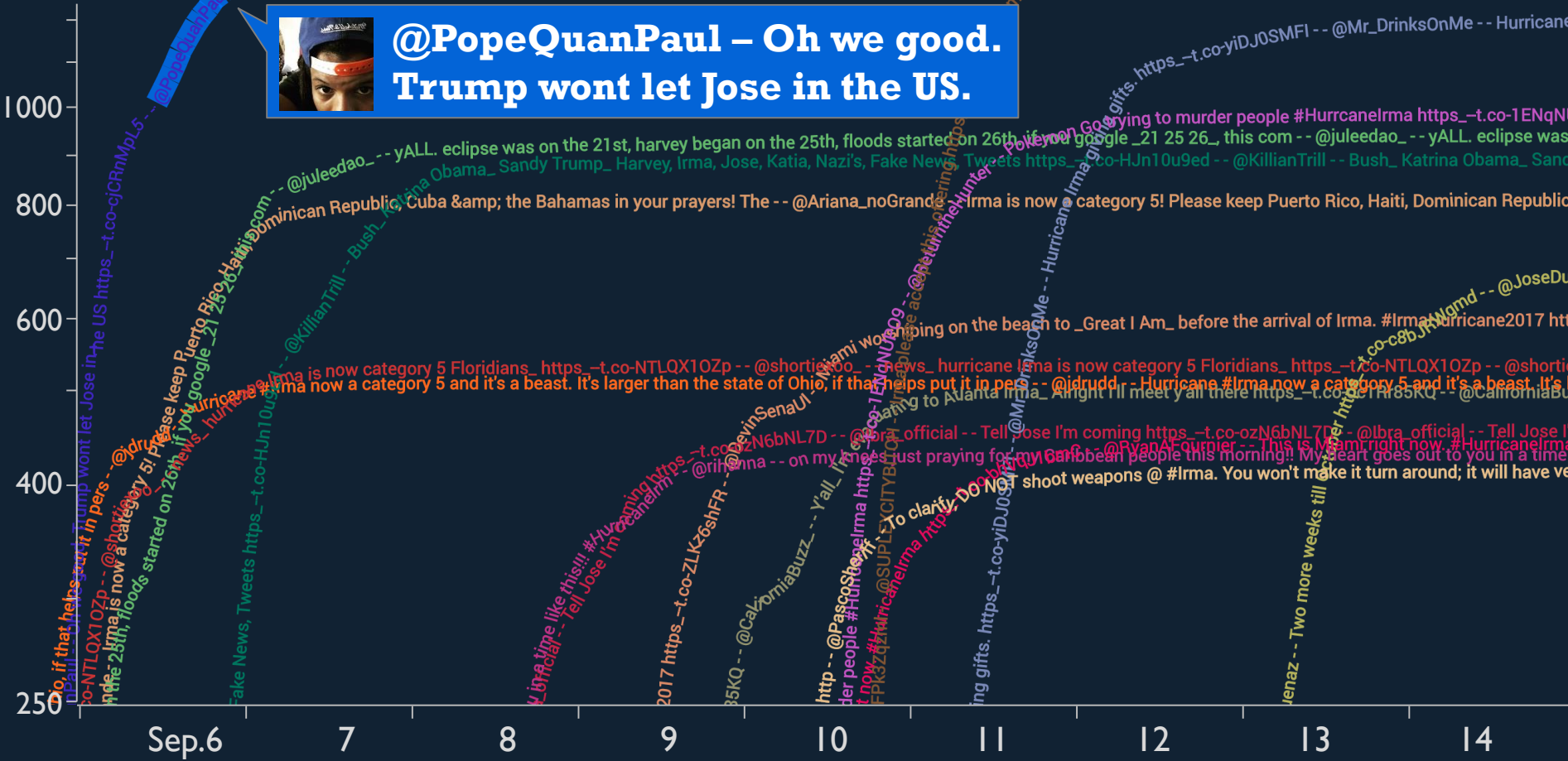
Top Hurricane Retweets

Sep. 6-14, 2017

Comedic flare...



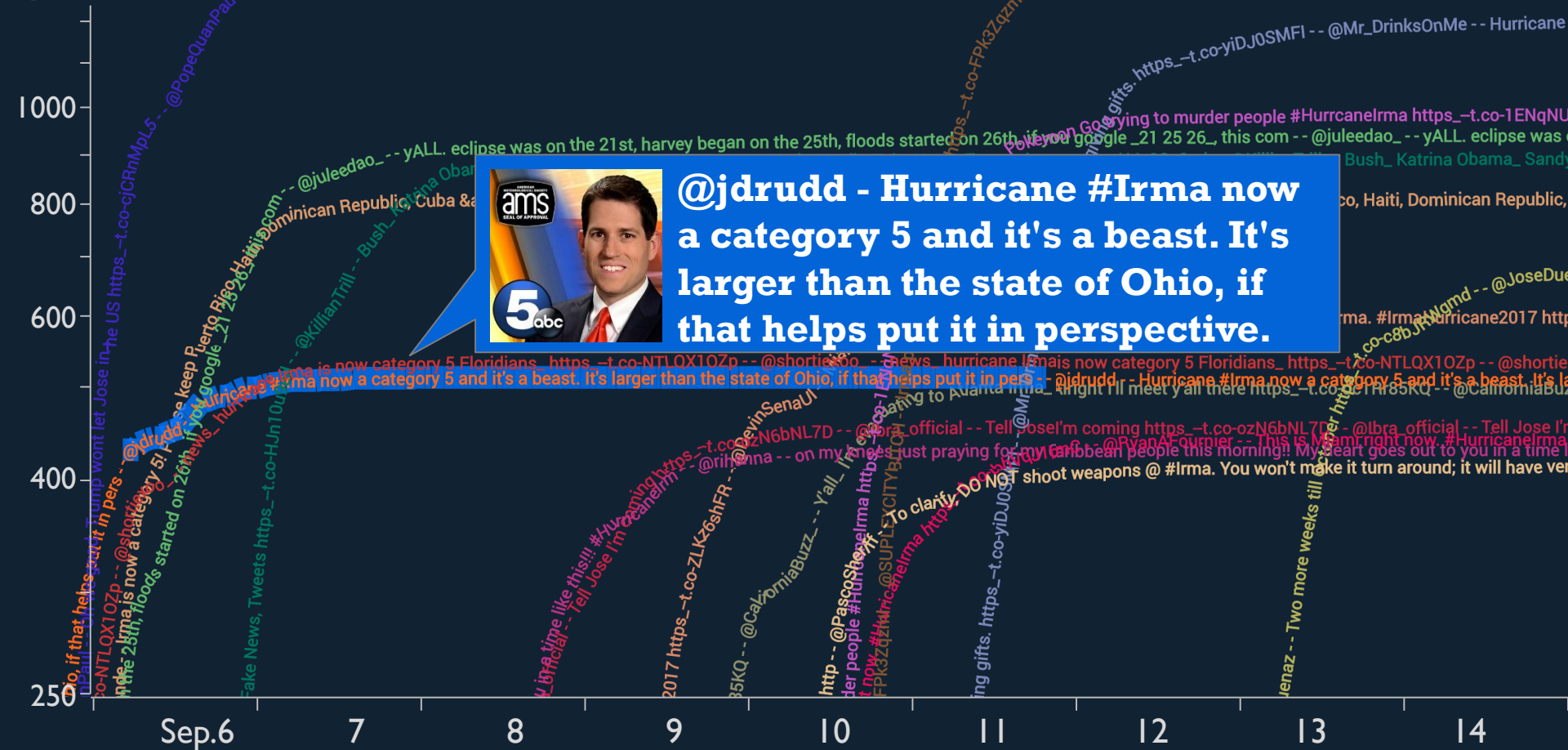
@PopeQuanPaul – Oh we good. Trump wont let Jose in the US.



Top Hurricane Retweets

Sep. 6-14, 2017

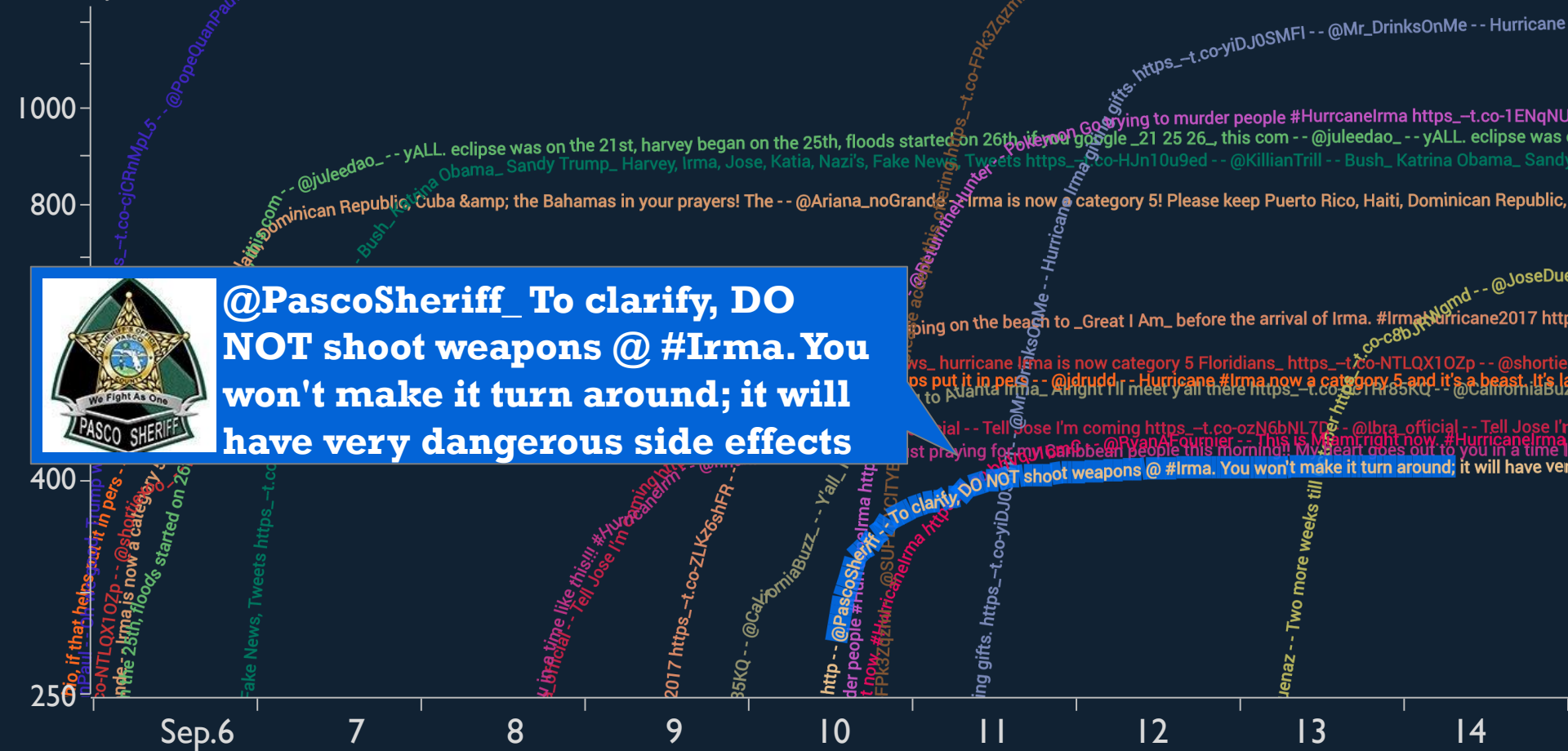
Important news...



Top Hurricane Retweets

Sep. 6-14, 2017

Public safety...



So What

Sticky is about a lot of things

- number retweets
- longevity
- reach across communities
- who's saying it
- what they're saying

And...

- geographic reach
- topics
- demographic segmentation
- message impact
- drift over time



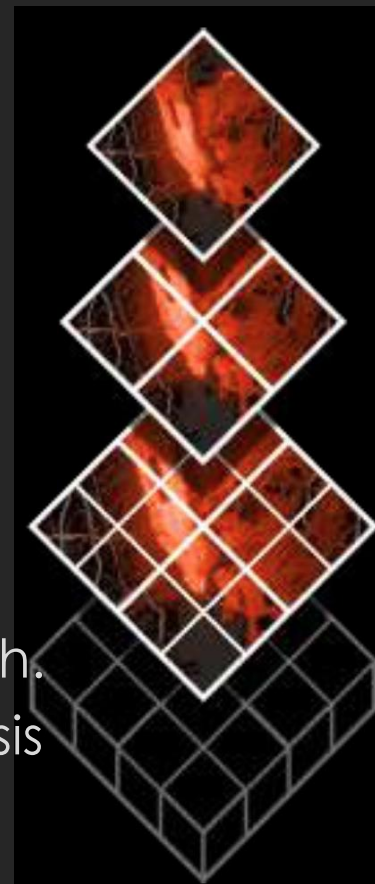
Salt

TILE-BASED VISUAL ANALYTICS

- Hierarchical data tiling using cluster computing.
- Interactive on-demand image tile generation.
- Layers of raw data and derivative analytics.

OPEN SOURCE

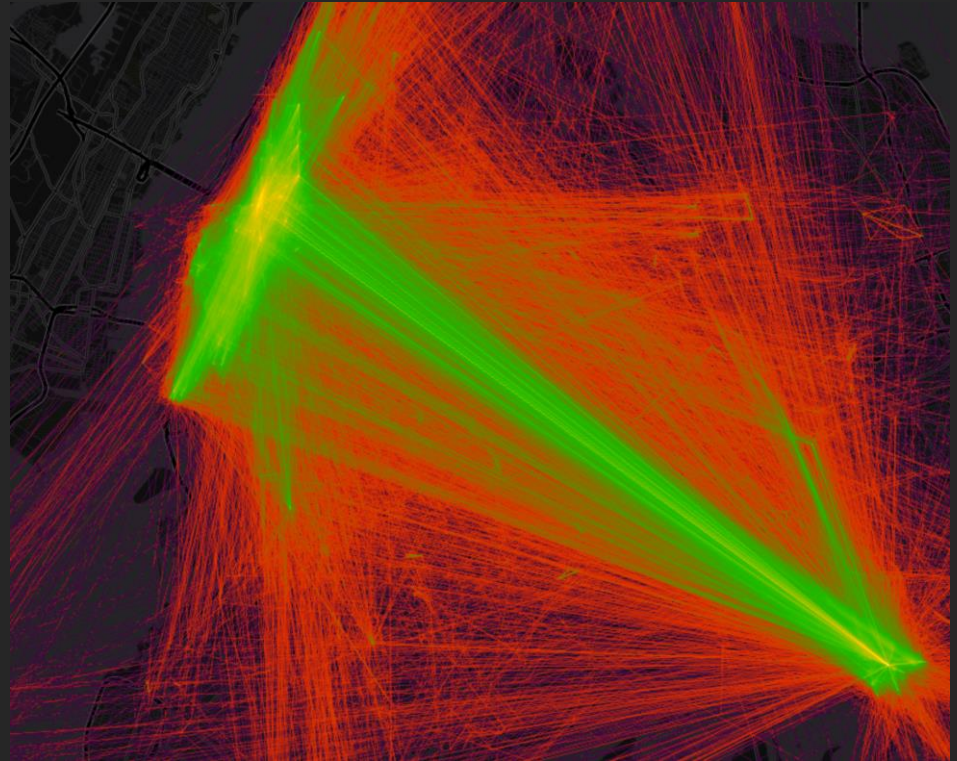
- Saltlib.com
- Built on Apache Spark, Hadoop and ElasticSearch.
 - Cross Plots, Geospatial, Time-series, Graph analysis



Geo-temporal Movement and Events View

People Trails

- By connecting users geo-coded social media posts ordered temporally aggregate “people trails” can be extracted and aggregated to see movement patterns
- User provides selection criteria to specify population segment to generate trails
 - Based upon topic or keywords, time and location or start with seed social media accounts and generate trails for community
 - E.g. generate trails for all people going through Kennedy Airport June 1 – June 20



Raw Data Drilldown

Longitude: -73.81° to -73.81°
Latitude: 40.70° to 40.70°
Total Tweets: 104

Top 10 user distribution (86 total)



Top 25 hashtag distribution (29 total)



Users

Hashtags

User: **nytokenblackguy**

Total Tweets: 5

Most Recent Tweets

translate

September 13, 2013

NYtokenblackguy

Even as a passenger I hate the Airtrain. (@ JFK
AirTrain - [Jamaica Station](#)) <http://t.co/uB81FesS7D>

8:36:00 am

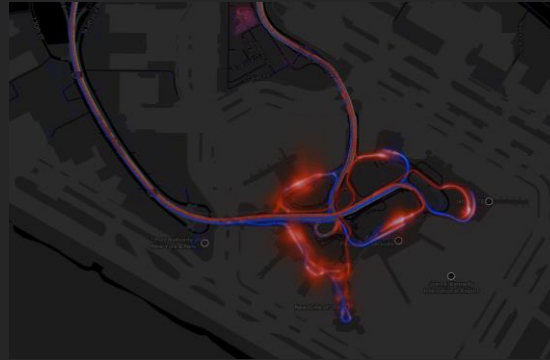


Longitude: -73.81° to -73.81°
Latitude: 40.70° to 40.70°

Geo-temporal Movement and Events View

Delta-plots and heatmaps

- Use heatmaps of activity in NYC taxi or social media to summarize overall activity and hot spots
- Delta plots show differences from average behavior or previous time periods
 - See increases or decreases in typical behavior for selected time period
- Supports identifying abnormal patterns and spatial extents
- Filtering by frequency allows focusing in on areas of largest activity or change, e.g. traffic problems during construction at the airport or change in usage during a street festival

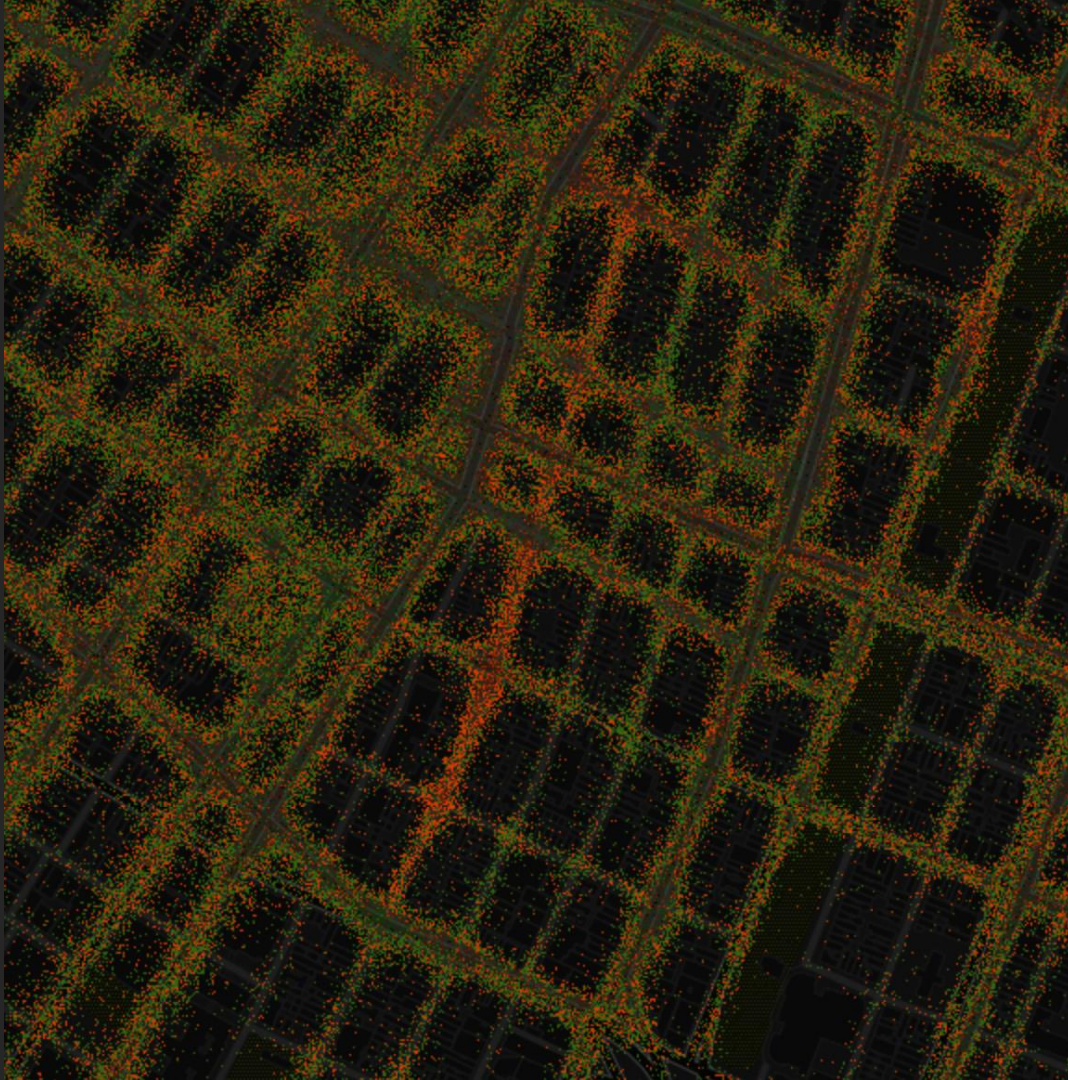


Polar color ramps show positive / negative difference from "normal" or previous time periods



Heatmaps with a perceptually correct color ramp summarize overall activity

The delta plot data is noisy, but still a red stripe is clearly visible in this plot indicating a big change such as an event occurring.



The most frequent hashtags (TFIDF) can be used to quickly characterize the event – in this case a street festival.

#sangennaro

Tweets by Time

Date	Tweet Volume
Aug 27	Low
Sep 10	Medium
Sep 24	High

Most Recent Tweets translate

September 22, 2013

JaroG4
Last day of the #SanGennaro festival. This place is delicious. From Houston to Canal, filled with tons...
<http://t.co/2YCTTh2QEo>
11:20:02 pm

DJWIKIINY
#SanGennaro Feast 2013 🇮🇹 @ Little Italy
<http://t.co/JoM0oYIbXo>
10:56:52 pm

DJWIKIINY
Twisted right now at the #SanGennaro feast 🇮🇹 In little Italy @flockent @ Little Italy <http://t.co/C6eromLydi>
10:08:34 pm

Cloud_ia
Working the #Robertas booth at the #sangennaro #festival #pizza #littleitaly #nyc @ Da Gennaro
<http://t.co/1POtStemi0>
9:52:18 pm

lifeofakidd
Enjoying the perfect Italian sausage and broccoli rabe

#bemelissa #profwomen
#aleakedfollowparty #ss14
#99school #engagenyc
#soho #follow #m240 #nyfw
#nyfw #cmmeetup #soho
#cronut #artcrank
#dianafollowparty #iamrarenyc
#contentconvo

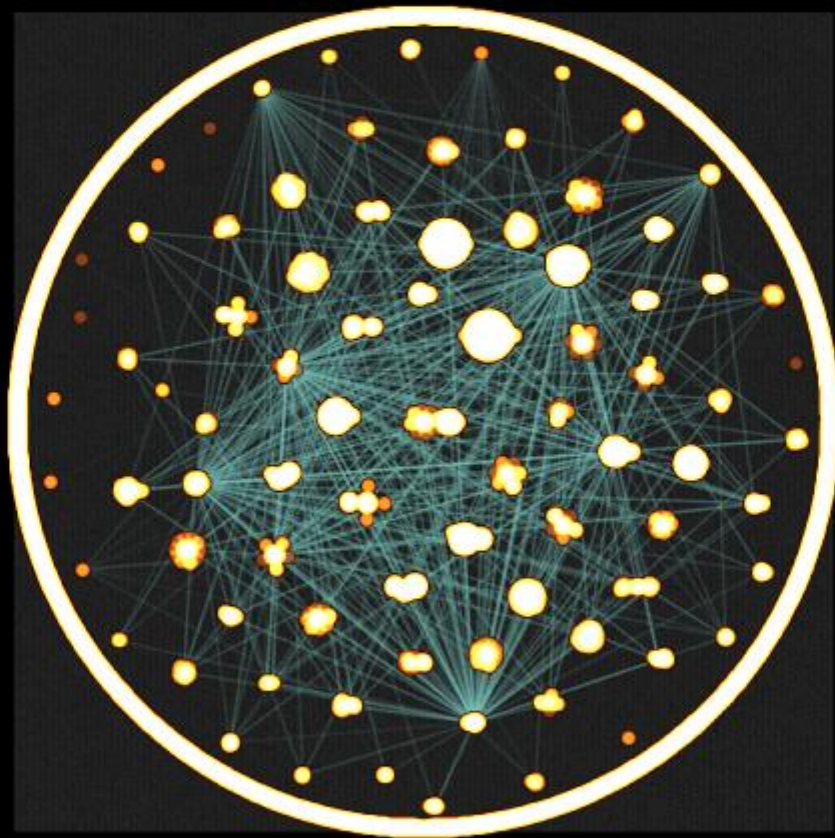
#americanapparel #thanksticketoak
#chinatown #brandgenius
#instavine #nyfw #yokoono
#soho #littleitaly #fb
#nyfw #sangennaro
#fashion #pixies
#indiecon #photography
#smalltalks #plasticonoband
#marketingjobs

#foodgasm #direnarmutlu
#1smx #theoriginals #festival #yummy
#gamified #sangennaro #travel
#nyfw #littleitaly
#chinatown #tourist #chinatown
#lebaron #nyc2013
#ahmetatakan
#libereneizócalo20smx
#foodstagram

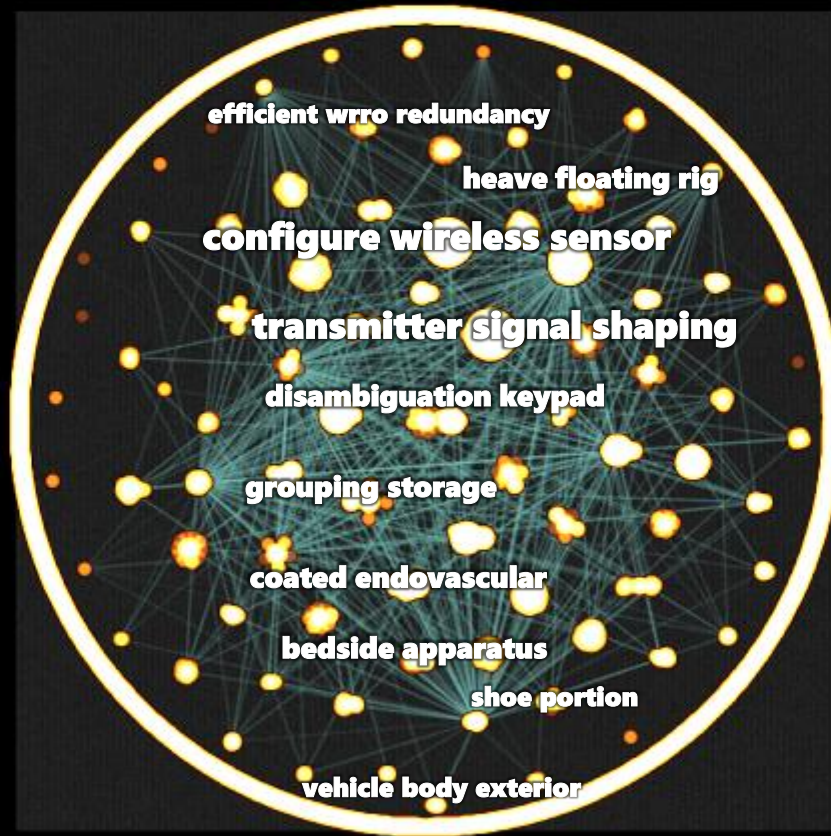
6. LITIGIOUS TOPICS

Intellectual Property

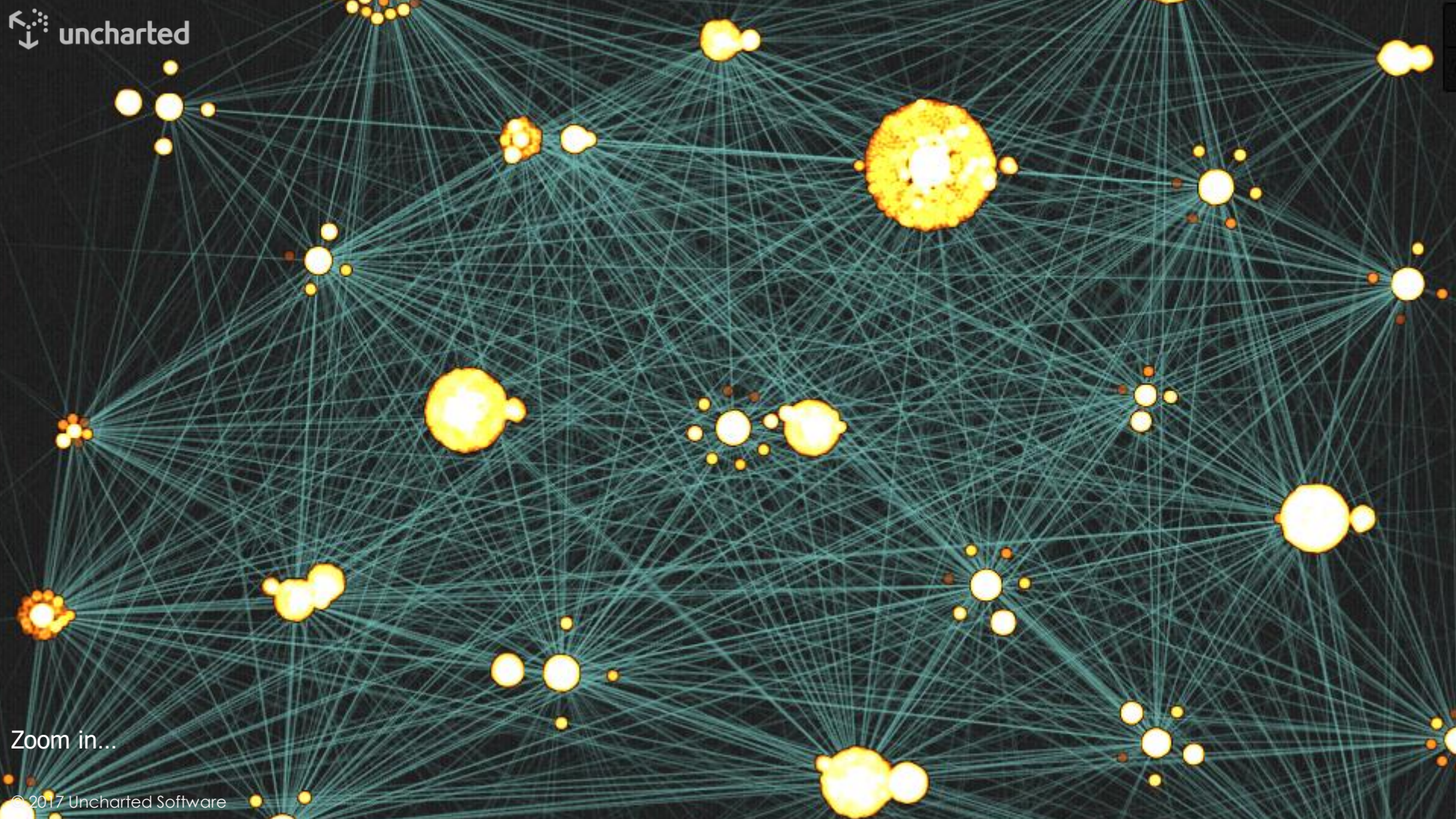
- Where are the hot spots?
 - What skills and expertise do we need?
- Where are the problem spots?
 - More diligent in crafting patent (more time, more resources)



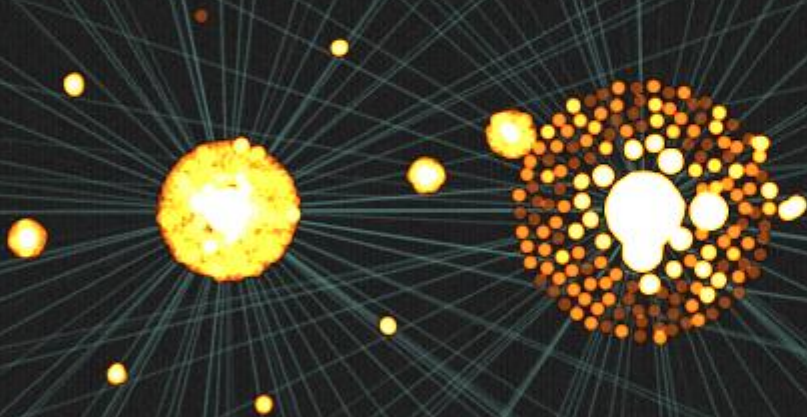
Plot showing every patent since 1982, grouped into communities based on each patent's citation of earlier relevant patents.



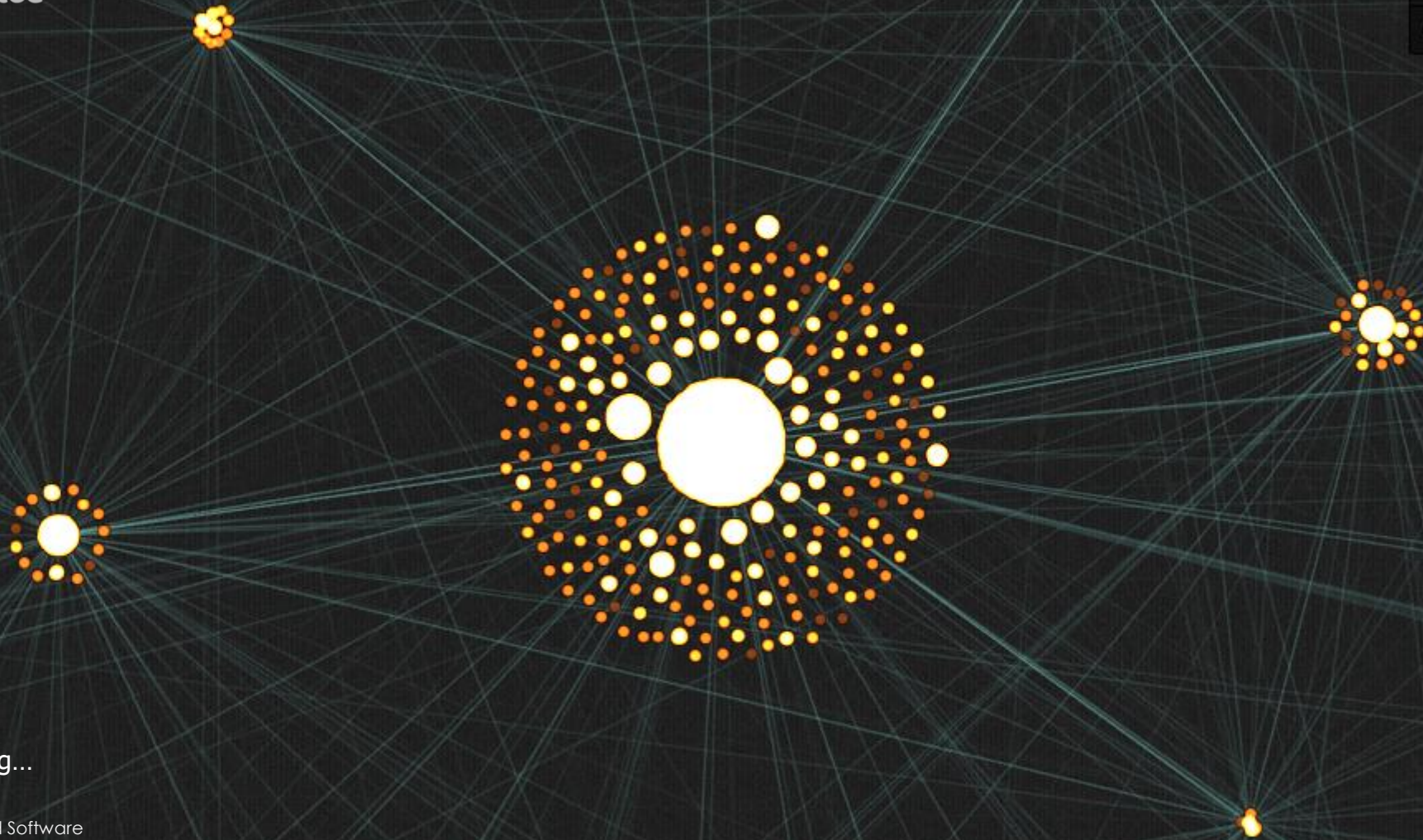
Labels, extracted using TFIDF, show the most unique terms on the most heavily cited patent in each community. These terms give an indication of the topics in each community.



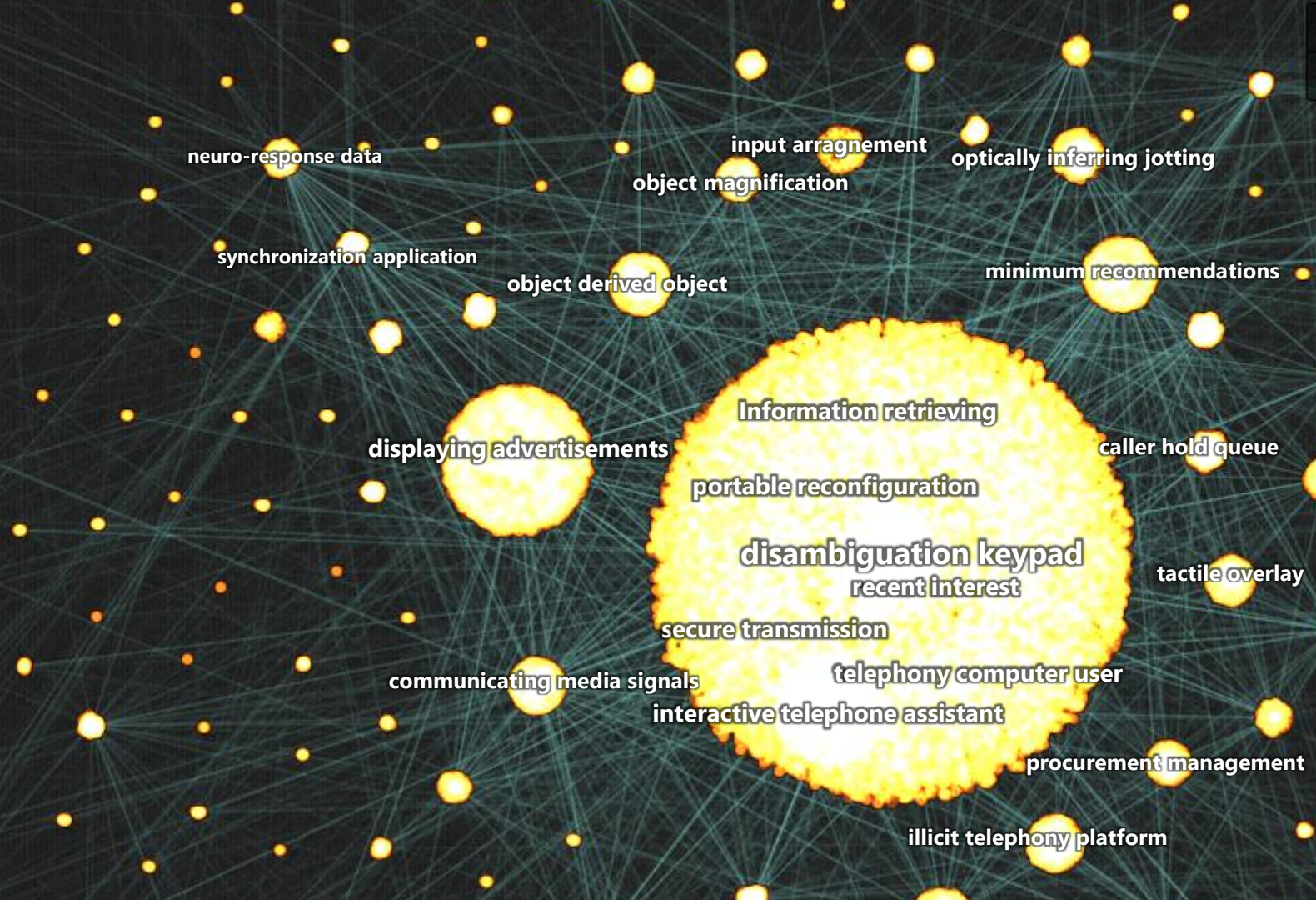
Zoom in...



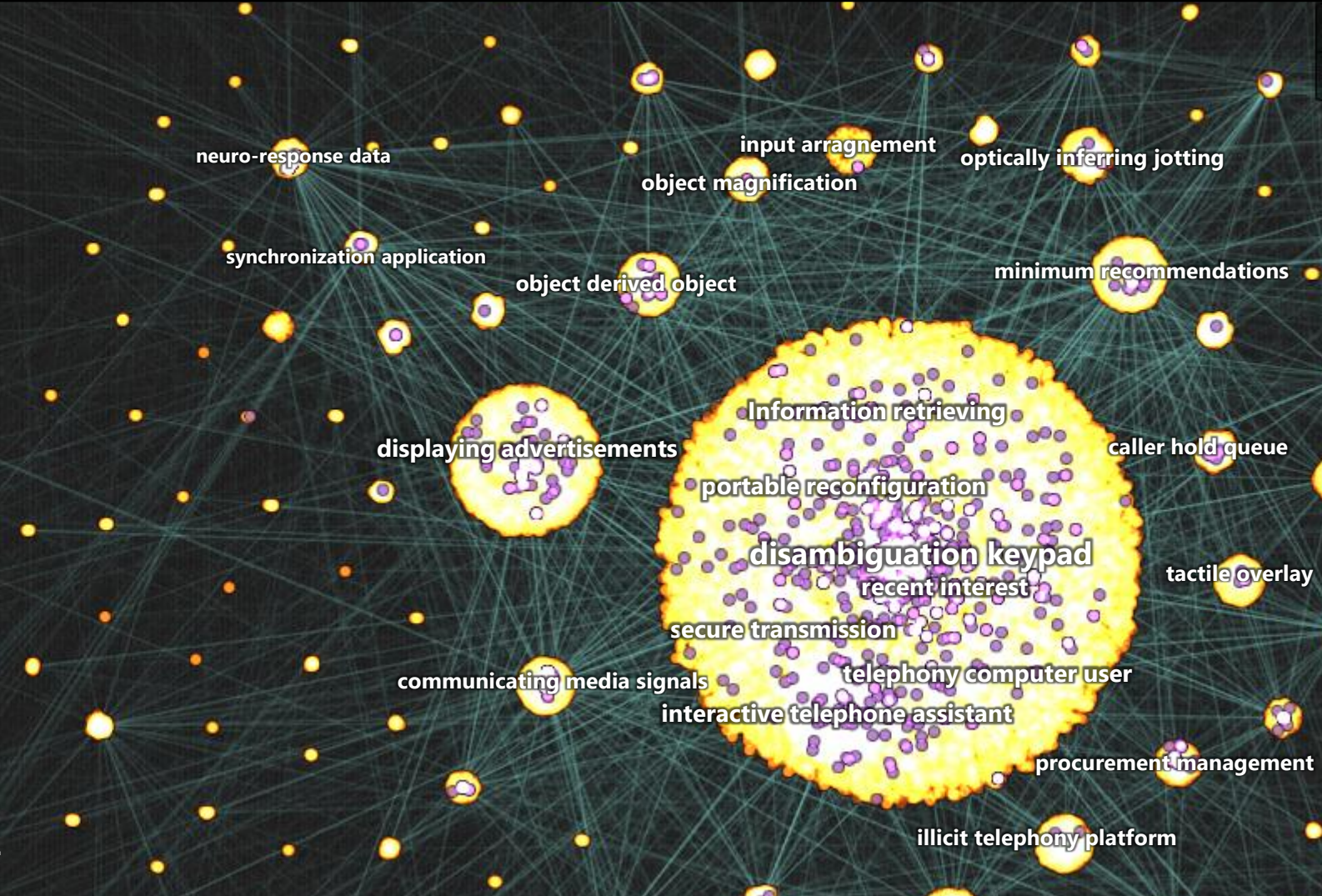
Keep zooming...



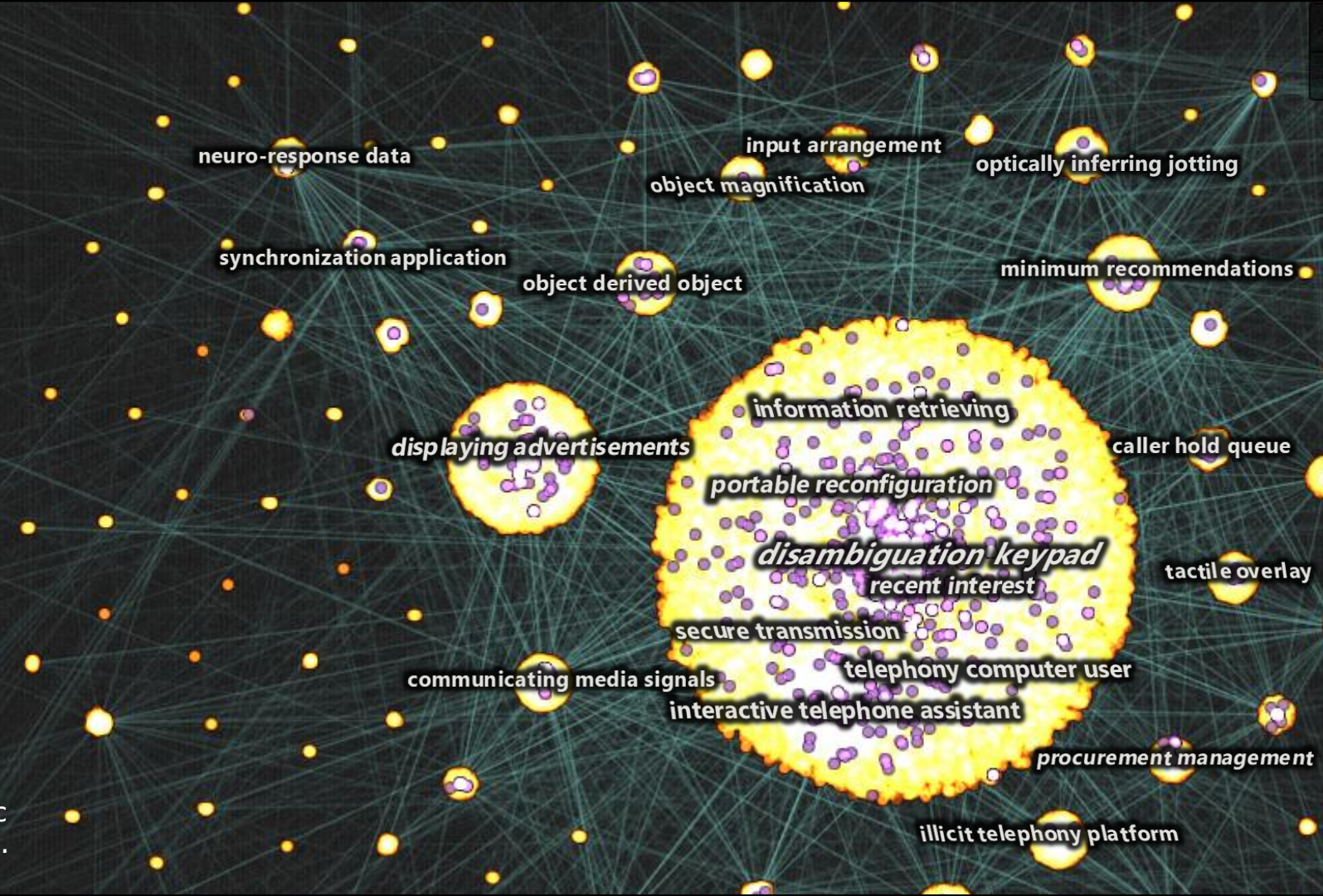
Keep zooming...



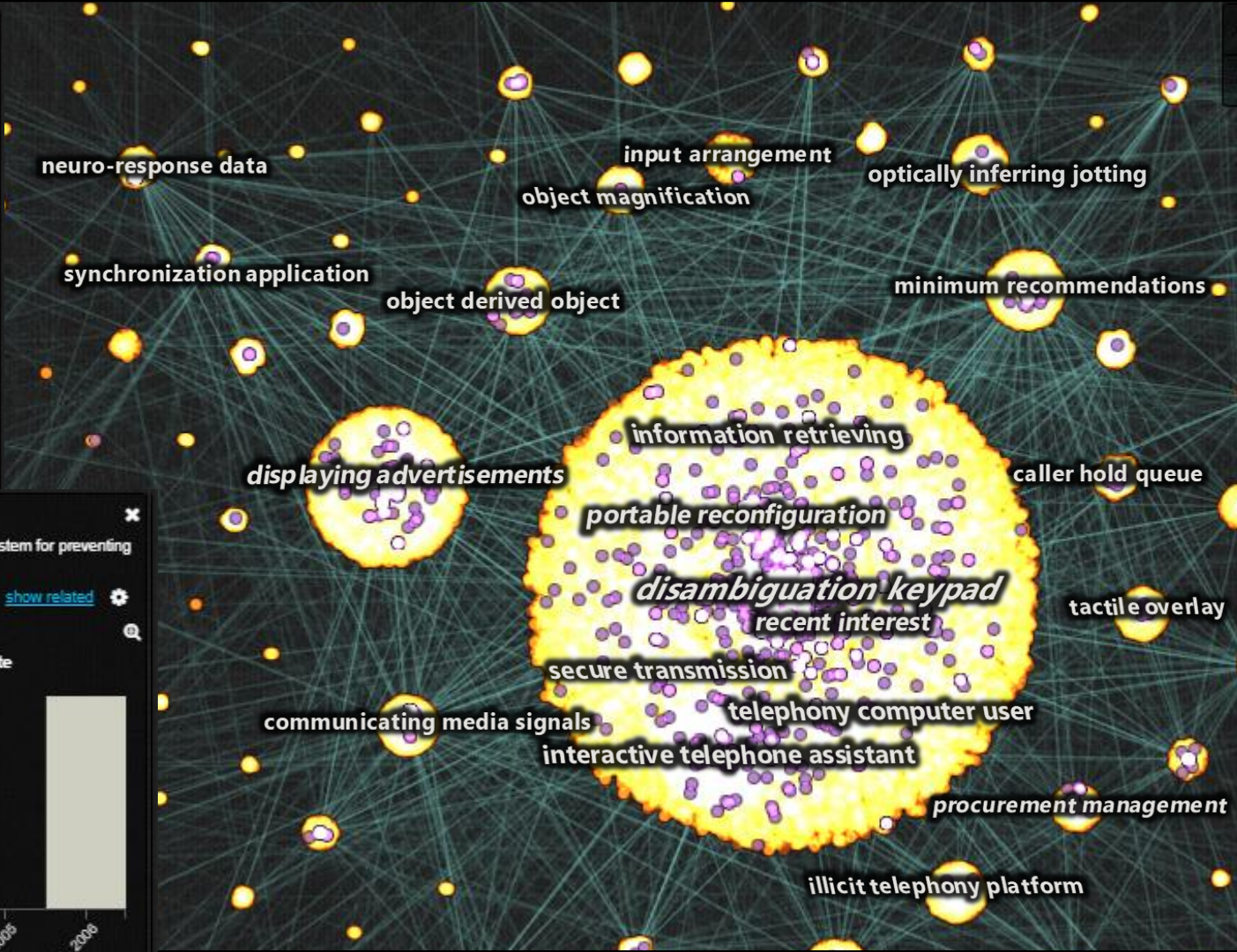
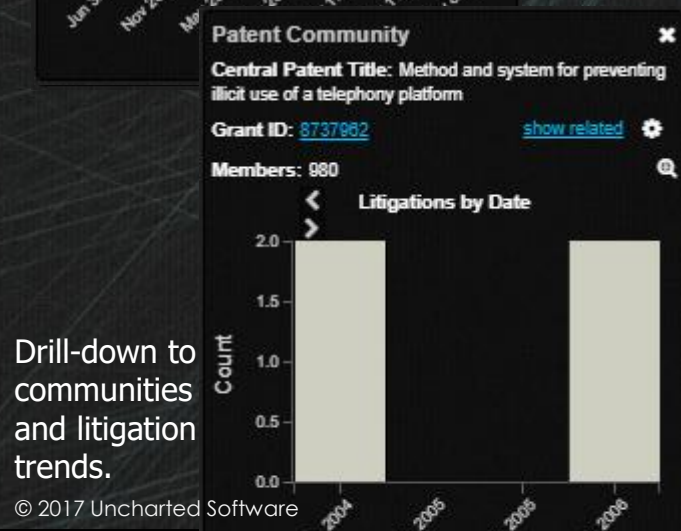
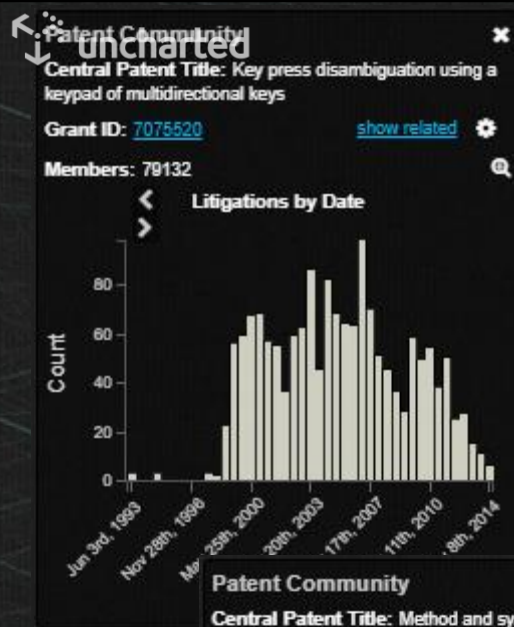
And focus on this community and sub-communicates. Labels indicate topics related to telephones.



Purple dots indicate patents with litigation.



Italic angle indicates litigation ratio in that community. Steep italic indicates high litigation, reverse italic indicates low litigation.



Drill-down to communities and litigation trends.



Patent Community

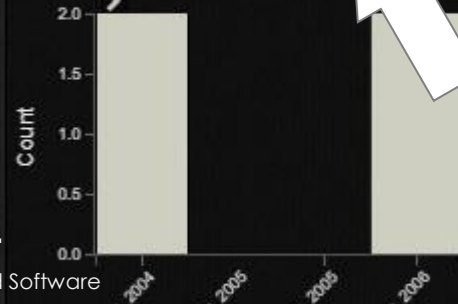
Central Patent Title: Method and system for preventing illicit use of a telephony platform

Grant ID: [8737962](#)

[show related](#)

Members: 980

Litigations by Date



See citations to other communities.

Method and system for preventing illicit use of a telephony platform

US 8737962 B2

ABSTRACT

A system and method for preventing illicit use of a telephony platform that includes enrolling a plurality of accounts on a telecommunications platform, wherein an account includes account configuration; at a fraud detection system of the telecommunications platform, receiving account usage data, wherein the usage data includes at least communication configuration data and billing configuration data of account configuration and further includes communication history of the plurality of accounts; calculating fraud scores of a set of fraud rules from the usage data, wherein at least a sub-set of the fraud rules include conditions of usage data patterns between at least two accounts; detecting when the fraud scores of an account satisfy a fraud threshold; and initiating an action response when a fraud score satisfies the fraud threshold.

Drill to patent publication.

IMAGES (9)

Publication number	US8737962 B2
Publication type	Grant
Application number	US 13/949,984
Publication date	27 May 2014
Filing date	24 Jul 2013
Priority date	24 Jul 2012

Also published as [US9270833](#), [5 More »](#)

Inventors [Adam Ballai](#), [Robert C. Hagemann, III](#), [Daniel Zarick](#)

Original Assignee [Twilio, Inc.](#)

Export Citation [BiBTeX](#), [EndNote](#), [RefMan](#)

[Patent Citations](#) (280), [Non-Patent Citations](#) (1), [Referenced by](#) (3), [Classifications](#) (11), [Legal Events](#) (1)

External Links: [USPTO](#), [USPTO Assignment](#), [Espacenet](#)

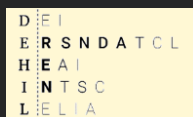
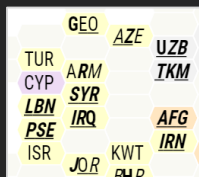
6. SO WHAT?

So What?

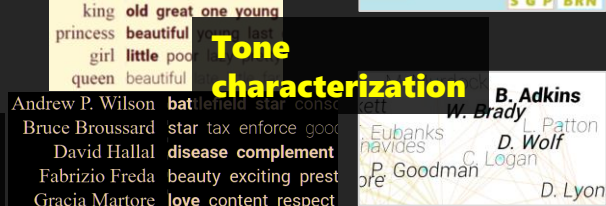
1. Visualization isn't just about numbers.
Text repositories may be 4x or more > numeric data.
2. The next big leap in analytics and visualization will be much deeper integration of *content* throughout the full *analytics + visualization* analysis cycle.
3. Real-world applications
 - Tone analysis of your (external) communications
 - Ongoing stickiness of (marketing) messages
 - Local topics and event detection
 - Litigious communities

Text Analytics and Visualization is Much Broader

Letters

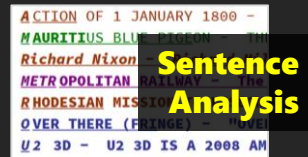
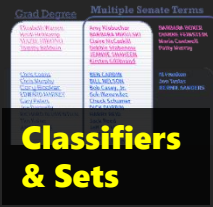


Words



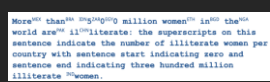
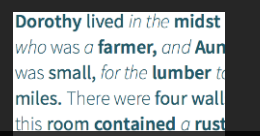
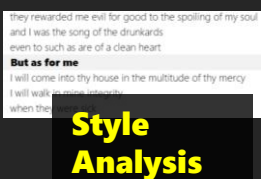
Phrases

Sentences



Paragraphs

Documents



What can you do?

- Understand what text data you have
- Define what business insights you want to extract

- Use open source tools for analytics and visualization
e.g. Saltlib, D3js, NLTK, Stanford Core NLP
- Incrementally experiment towards a solution
- Productionize the parts that are successful

THANK YOU

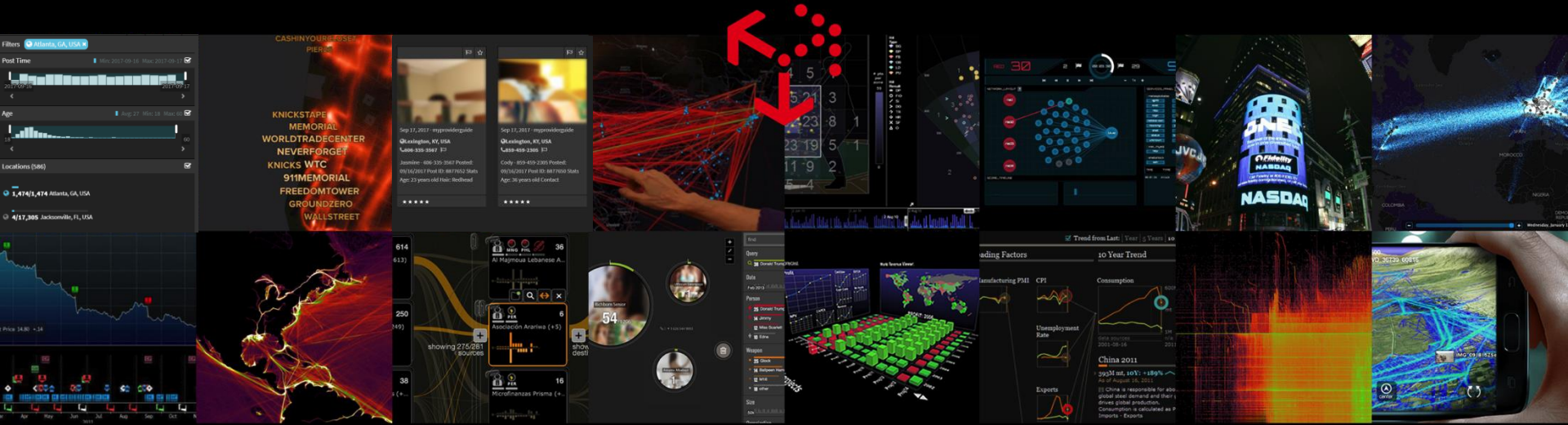
Richard Brath

rbrath@unchartedsoftware.com
@rkbrath
richardbrath.wordpress.com

Scott Langevin

slangevin@unchartedsoftware.com
@slangevi
blog.uncharted.software

uncharted



About Uncharted: Profoundly visual software for improved awareness, analysis and decision-making, used by hundreds of thousands in government, law enforcement, finance, and more.