

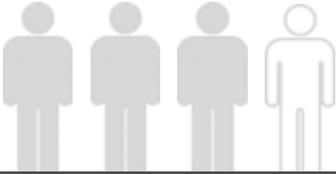
BIG DATA

Jiří Karpeta
ředitel vývojářské divize

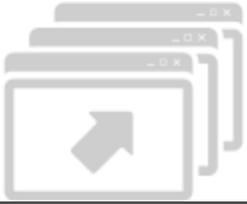
Microsoft ČR

Transformational trends

Millennials will make up **75%** of the American workforce by **2025**



Over **80%** of new apps will be distributed or deployed on clouds in **2012**.



Digital content will grow to **2.7ZB in 2012**, up 48% from 2011, rocketing toward **8ZB by 2015**.



By **2016**, smartphones and tablets will put power in the pockets of **a billion** global consumers.

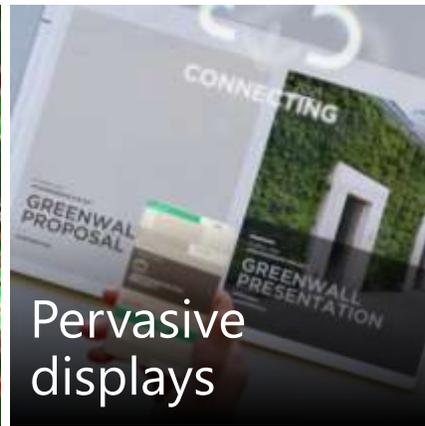
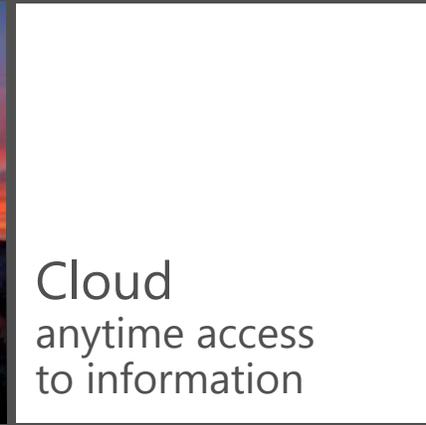


The world's **mobile worker** population will reach

1.3 billion

—over **37%** of the total workforce—by **2015**.

IT megatrends



65% of companies are deploying at least one **social software tool**.



70% of businesses are either using or investigating **cloud computing solutions**

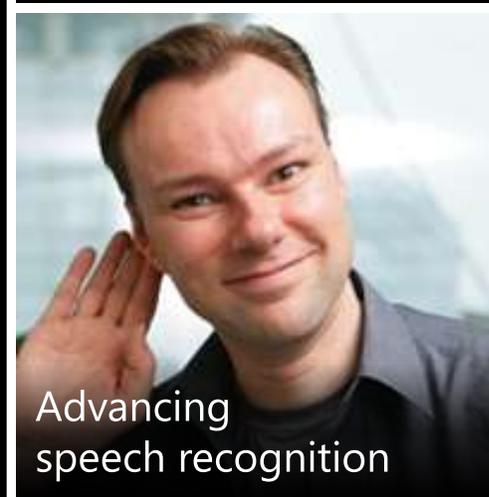
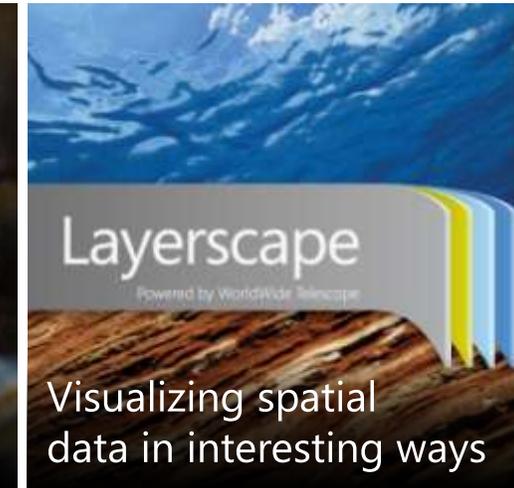


80% growth of unstructured data is predicted over the **next five years**.



We live in an amazing time!

Microsoft Research projects in action



Big Data?

Data volume
doubles every
40 months

Daily
production of
bytes
 2.5×10^{18}



Connected
systems
Internet
usage



Hard to store

Hard to process in real time

Models are more important

Every industry affected

Social disciplines

Research

And the Oscar Goes

to ...

David M. Rothschild, PhD

I was very accurate with **19 of 24 categories correct...**

**PredictWise's 2013 Oscar Likelihood of Victory at 4 PM ET on Sunday, February 24 (Oscar Night!)
(The actual winners are highlighted in yellow or orange)**

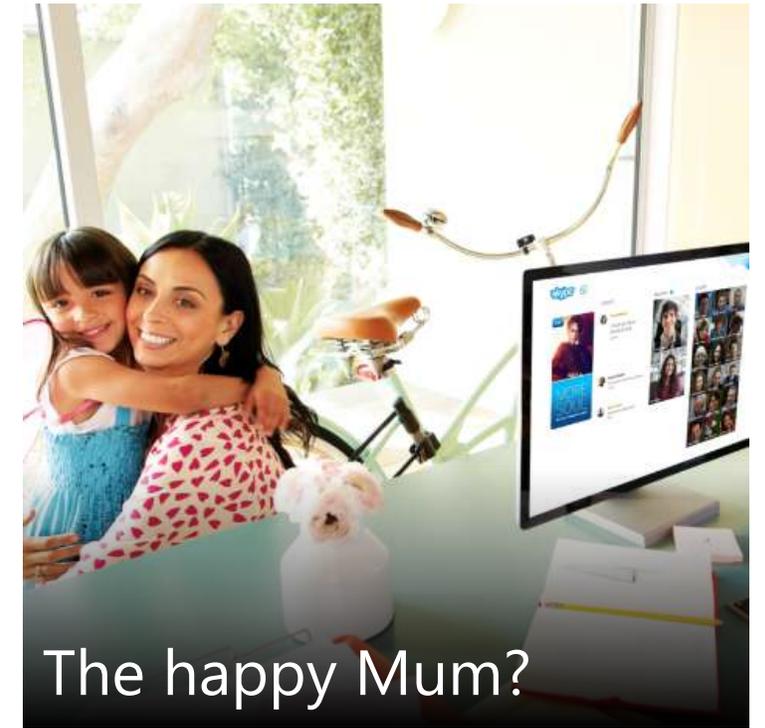
Best Director			Best Adapted Screenplay		Best Cinematography	
Steven Spielberg	Lincoln	76%	Argo	68%	Life of Pi	95%
Ang Lee	Life of Pi	22%	Lincoln	26%	Skyfall	5%
David O Russell	Silver Linings Playbook	1%	Silver Linings Playbook	5%	Lincoln	0%
Michael Haneke	Amour	0%	Life of Pi	1%	Anna Karenina	0%
Benh Zeitlin	Beasts of the Southern Wild	0%	Beasts of the Southern Wild	0%	Django Unchained	0%
Best Actor			Best Original Screenplay		Best Costume Design	
Daniel Day-Lewis	Lincoln	99%	Django Unchained	52%	Anna Karenina	87%
Hugh Jackman	Les Miserables	1%	Zero Dark Thirty	27%	Les Miserables	13%
Joaquin Phoenix	The Master	0%	Amour	20%	Mirror Mirror	0%
Bradley Cooper	Silver Linings Playbook	0%	Moonrise Kingdom	0%	Lincoln	0%
Denzel Washington	Flight	0%	Flight	0%	Snow White and the Huntsman	0%
Best Actress			Best Foreign Language Film		Best Makeup and Hairstyling	
Jennifer Lawrence	Silver Linings Playbook	74%	Amour	99%	The Hobbit: An Unexpected Journey	68%
Emmanuelle Riva	Amour	23%	A Royal Affair	1%	Les Miserables	32%
Jessica Chastain	Zero Dark Thirty	3%	Kon-Tiki	0%	Hitchcock	1%
Naomi Watts	The Impossible	0%	No	0%		
Quvenzhané Wallis	Beasts of the Southern Wild	0%	War Witch	0%		
Best Supporting Actor			Best Animated Feature Film		Best Film Editing	
Tommy Lee Jones	Lincoln	44%	Brave	47%	Argo	84%
Christoph Waltz	Django Unchained	40%	Wreck It Ralph	46%	Zero Dark Thirty	16%
Robert De Niro	Silver Linings Playbook	14%	Frankenweenie	6%	Life of Pi	0%
Philip Seymour Hoffman	The Master	2%	ParaNorman	1%	Lincoln	0%
Alan Arkin	Argo	0%	The Pirates Band of Misfits	0%	Silver Linings Playbook	0%
Best Supporting Actress			Best Animated Short Film		Best Production Design	
Anne Hathaway	Les Miserables	99%	Paperman	70%	Les Miserables	46%
Sally Field	Lincoln	1%	Adam and Dog	7%	Anna Karenina	39%
Helen Hunt	The Sessions	0%	Fresh Guacamole	7%	Lincoln	9%
Jacki Weaver	Silver Linings Playbook	0%	Head Over Heels	7%	Life of Pi	6%

- Big Data on small dataset
- Category correlation
- User input
- Social media
- Box office returns
- Real time
- Universal model

From Demographics to Psychometrics....

Yoram Bachrach, PhD

- People's behaviour online in terms of psychometric measurements
- Big-5 personality traits
 - openness, conscientiousness, extraversion, agreeableness and neuroticism
 - general intelligence
 - satisfaction-with-life
- Patterns of Facebook usage
- Website preferences
- Query logs
- Facebook Likes
- Model to predict users behaviours, preferences or characteristics.

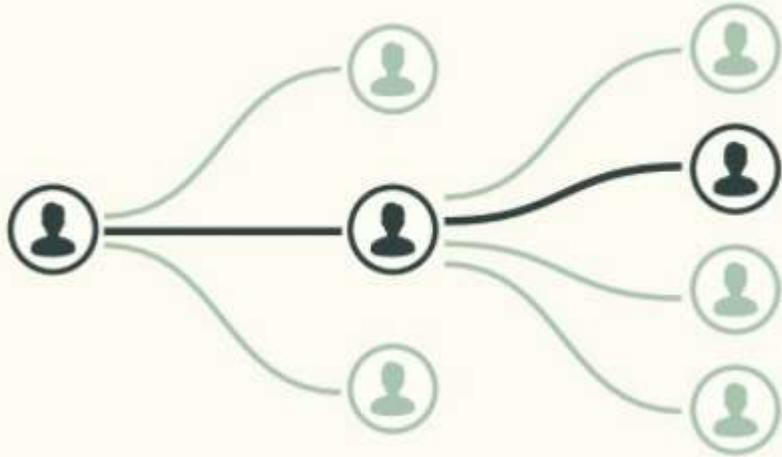


The happy Mum?

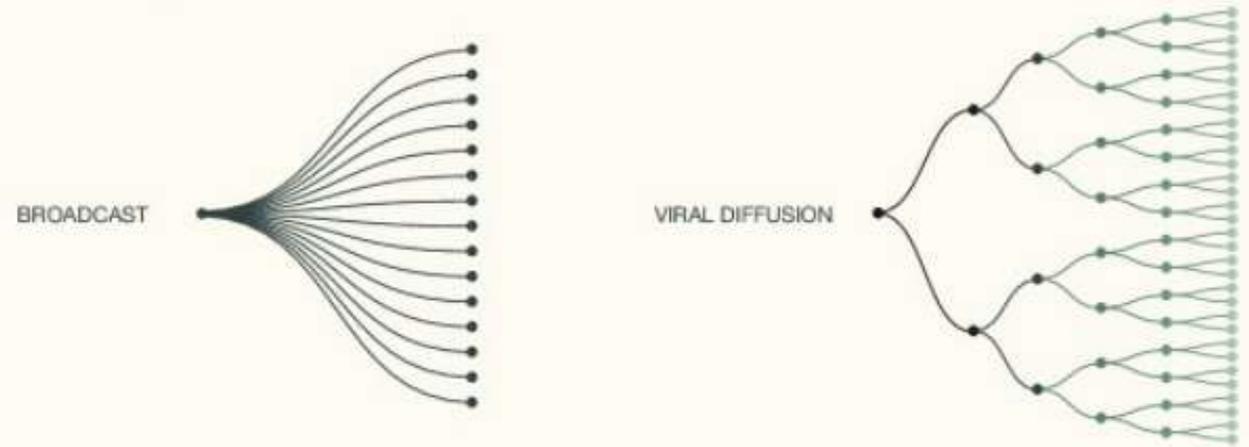
Identifying & Visualizing Viral Content....

Jake Hofman , PhD

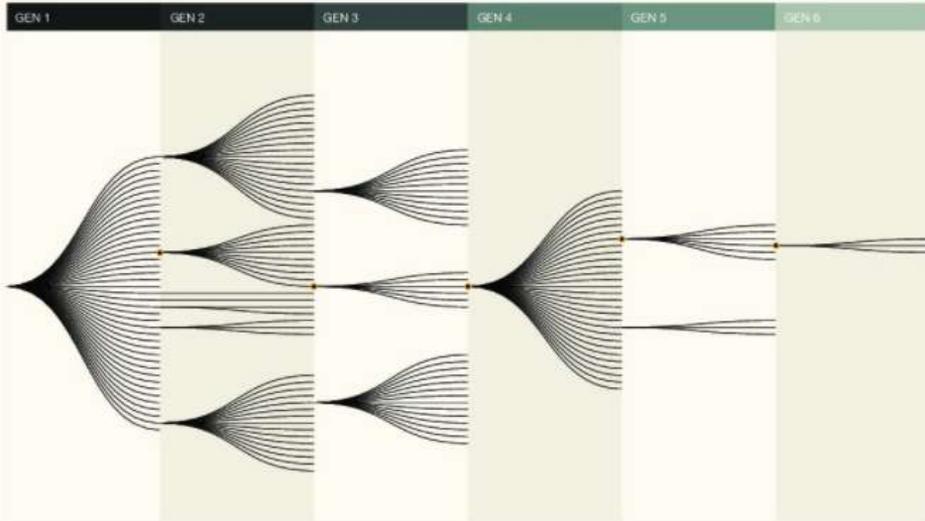
INFORMATION DIFFUSION



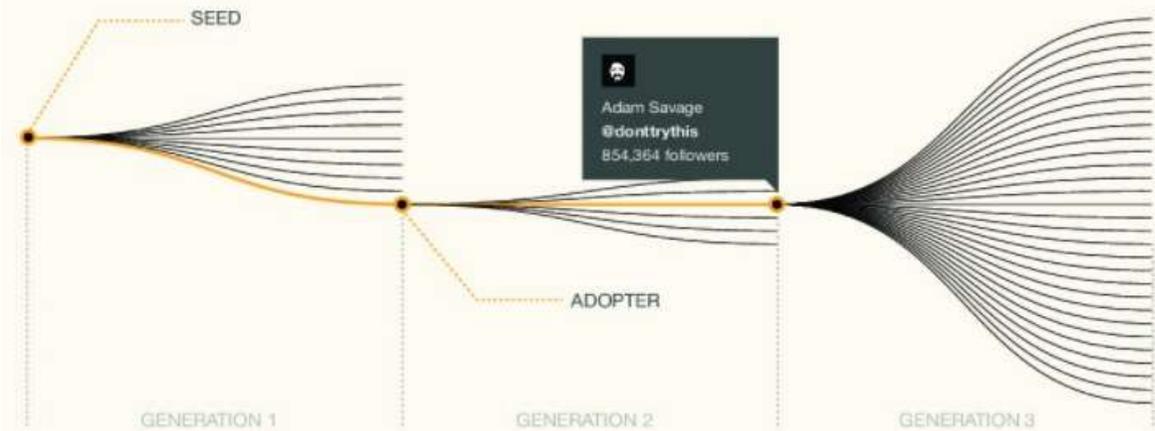
A schematic depiction of broadcast versus viral diffusion, where nodes represent individual adoptions and edges indicate who adopted from whom



TREES

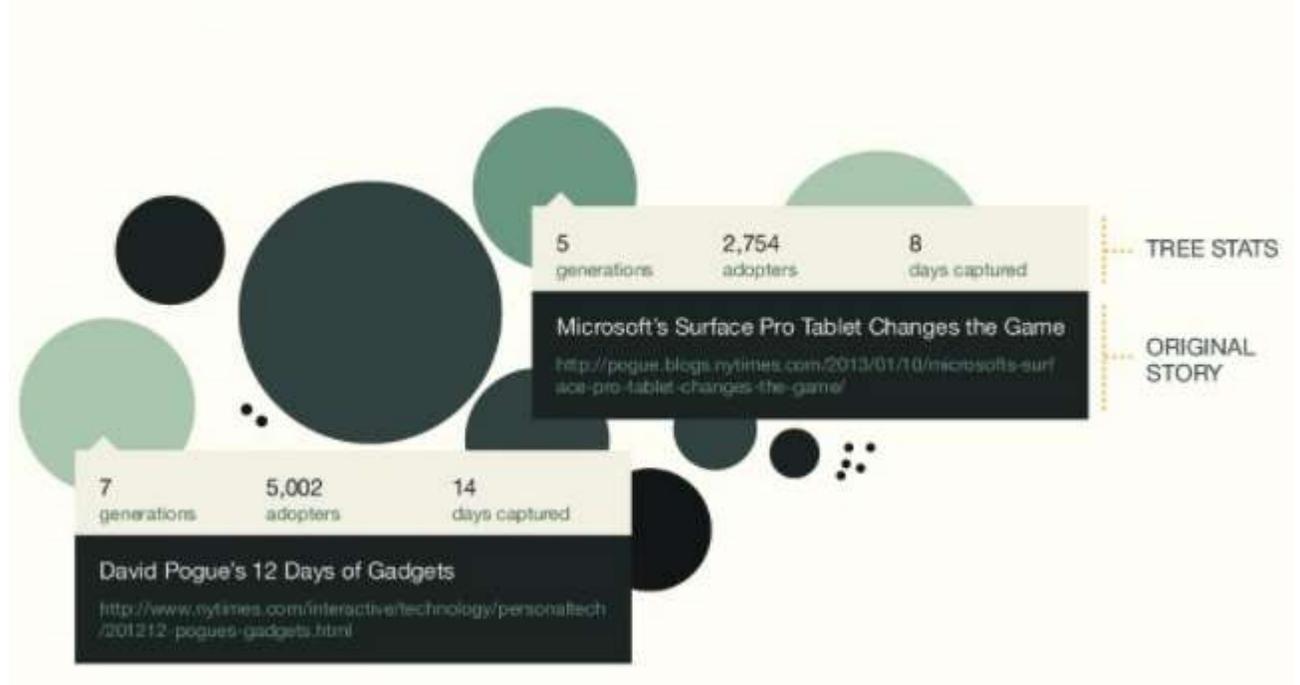


Tree view provides an interactive visualization depicting the spread of a particular story from one person to the next, starting from an initial seed user and propagating through multiple generations.



PEOPLE

People view offers a different perspective, showing details for all of the popular stories an individual has participated in.



METHODS & RESULTS

Examine **1.4 billion** tweets over one year

Trace the diffusion of each story over **1.0 billion** distinct trees

Most trees are small, with fewer than **2** adoptions on average

