

Search and the 'Net in 2017

Michael Hunter

Reference Librarian (part-time)
Hobart and William Smith Colleges

**For Rochester Regional Library Council
Member Libraries' Staff**

Sponsored by the

Rochester Regional Library Council

Supported by Regional Bibliographic Databases and Resources Sharing (RBDB) funds
granted by the

New York State Library 2017

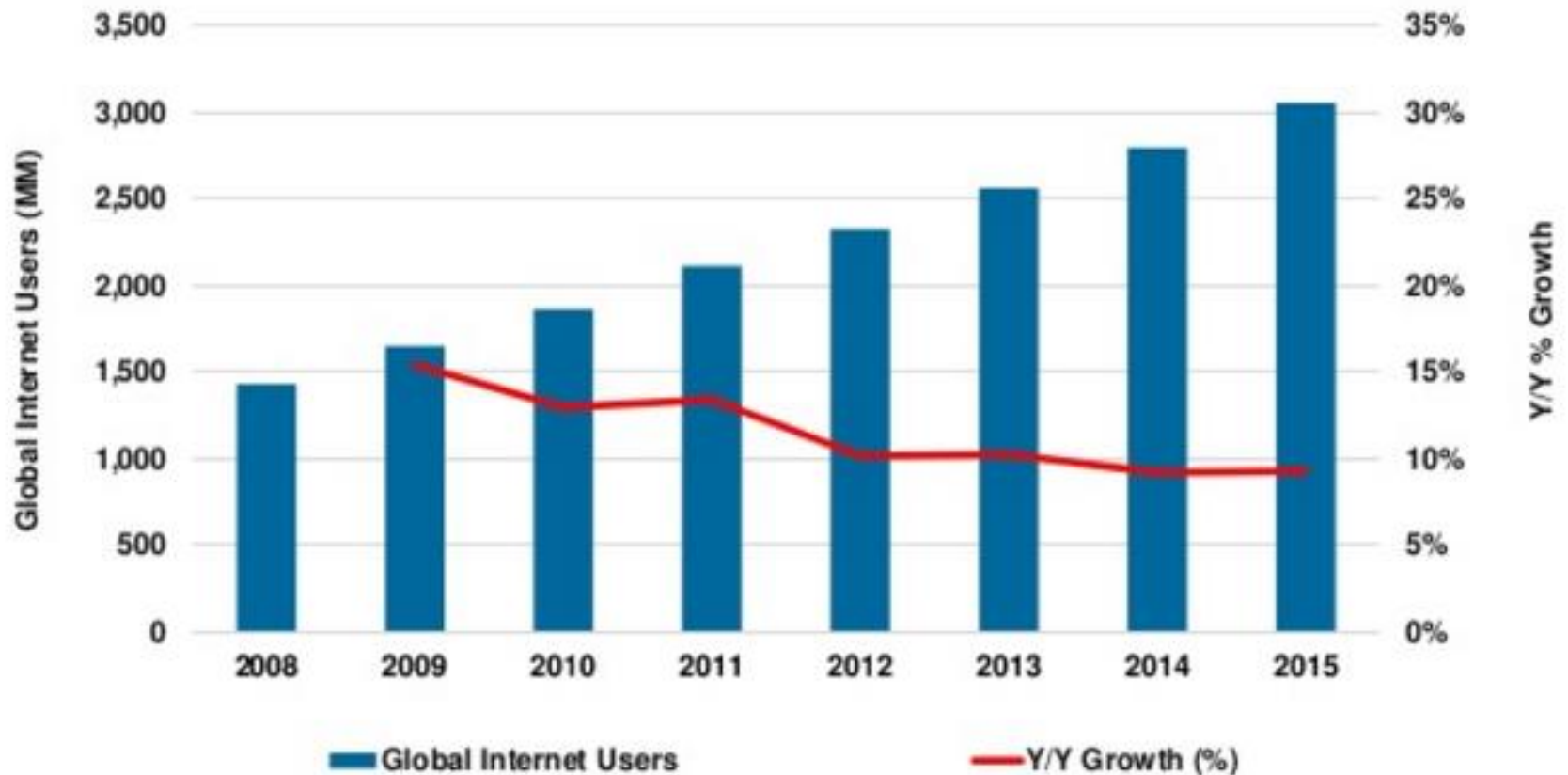
For today.....

- Landscape of Search
- Artificial Intelligence and Bias
- New Services
- The Social Web- Beyond the Personal
- The Year at Google
- Latest From Other Established Services
- Web Search in the Future

Internet Trends of 2016-Mary Meeker

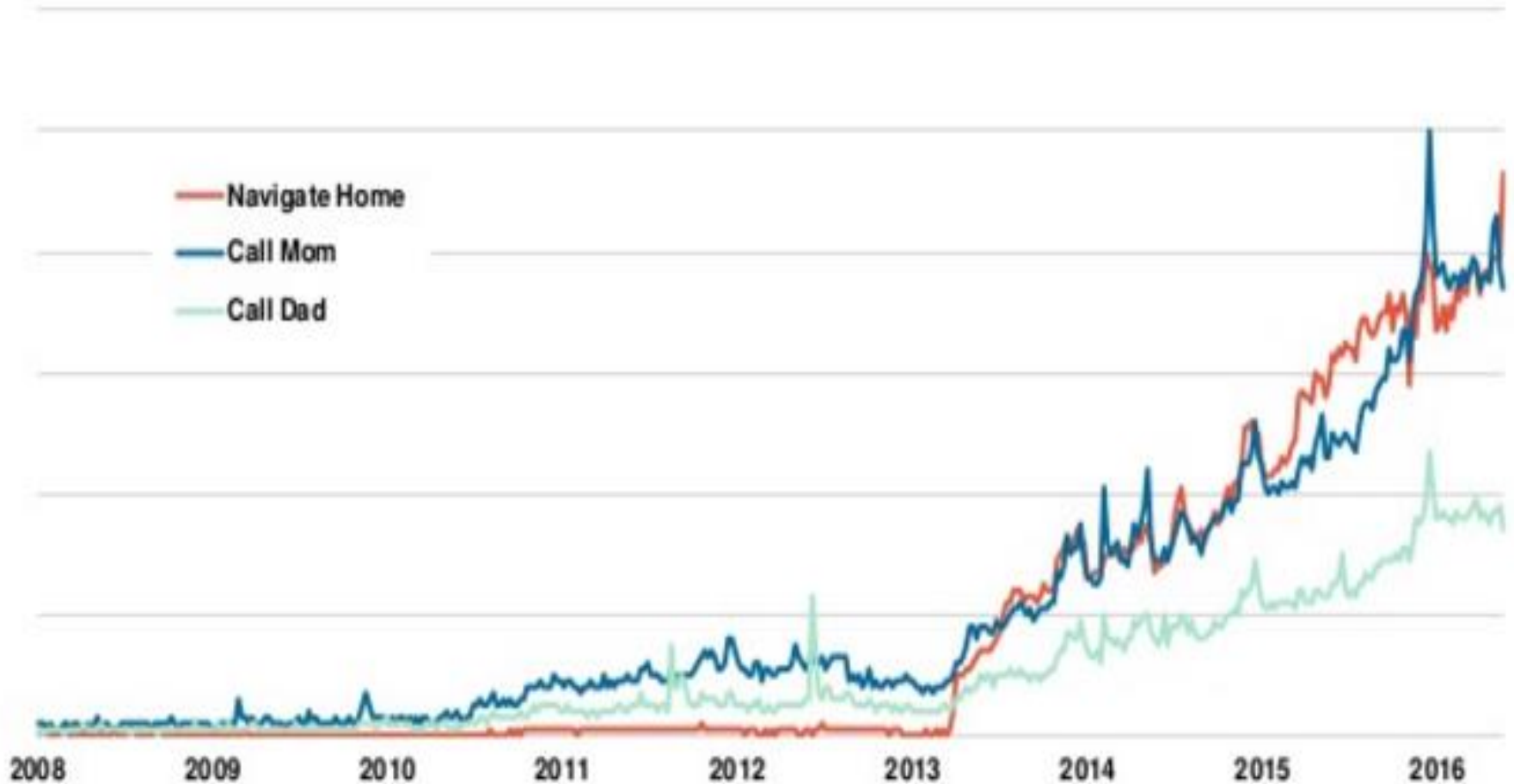
www.kpcb.com/internet-trends

Global Internet Users, 2008 – 2015



Google Voice Search Queries

Google Trends, Worldwide, 2008 – 2016



Attributes – Millennials vs. Gen Z

Millennials

Tech Savvy: 2 screens at once

Communicate with text

Curators and Sharers

Now-focused

Optimists

Want to be discovered

vs

Gen Z

Tech Innate: 5 screens at once

Communicate with images

Creators and Collaborators

Future-focused

Realists

Want to work for success

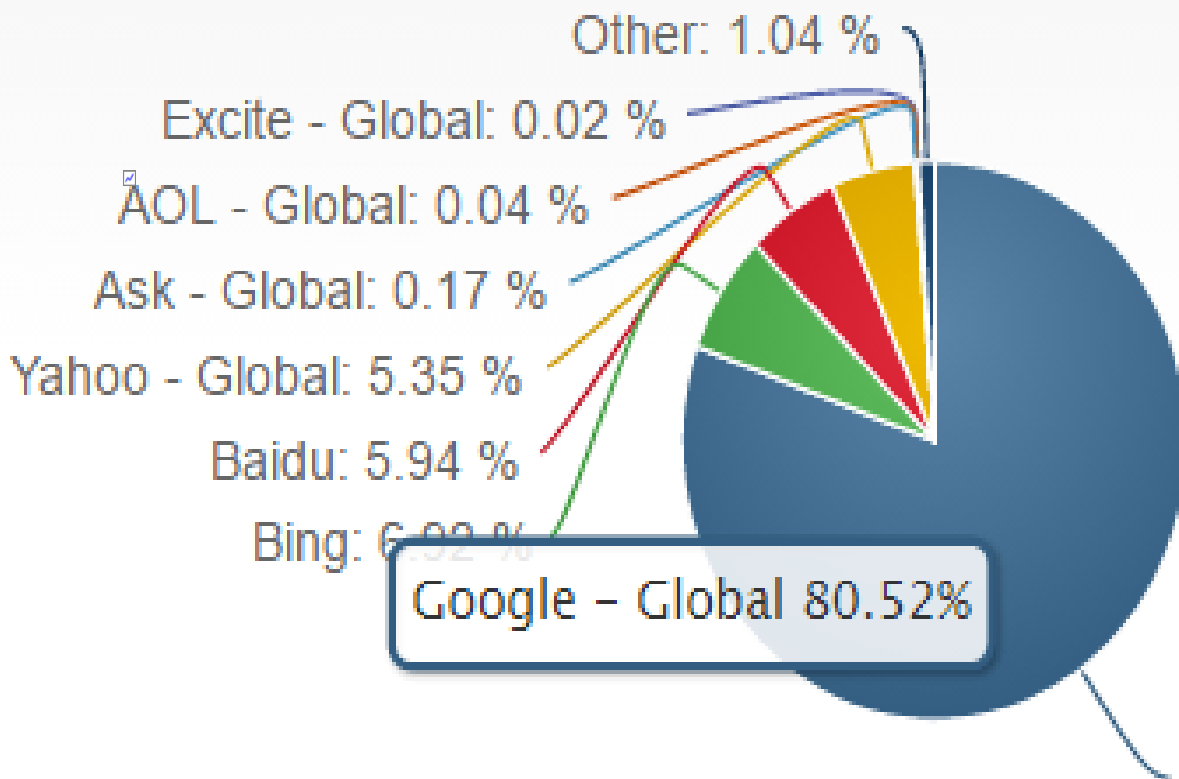
General Web Search in 2017

Major players crawling the Web

- **Google**
- **Bing (aka Yahoo! aka AOL)**
- **Gigablast**
- **DuckDuckGo**
- **Baidu**
- **Yandex**

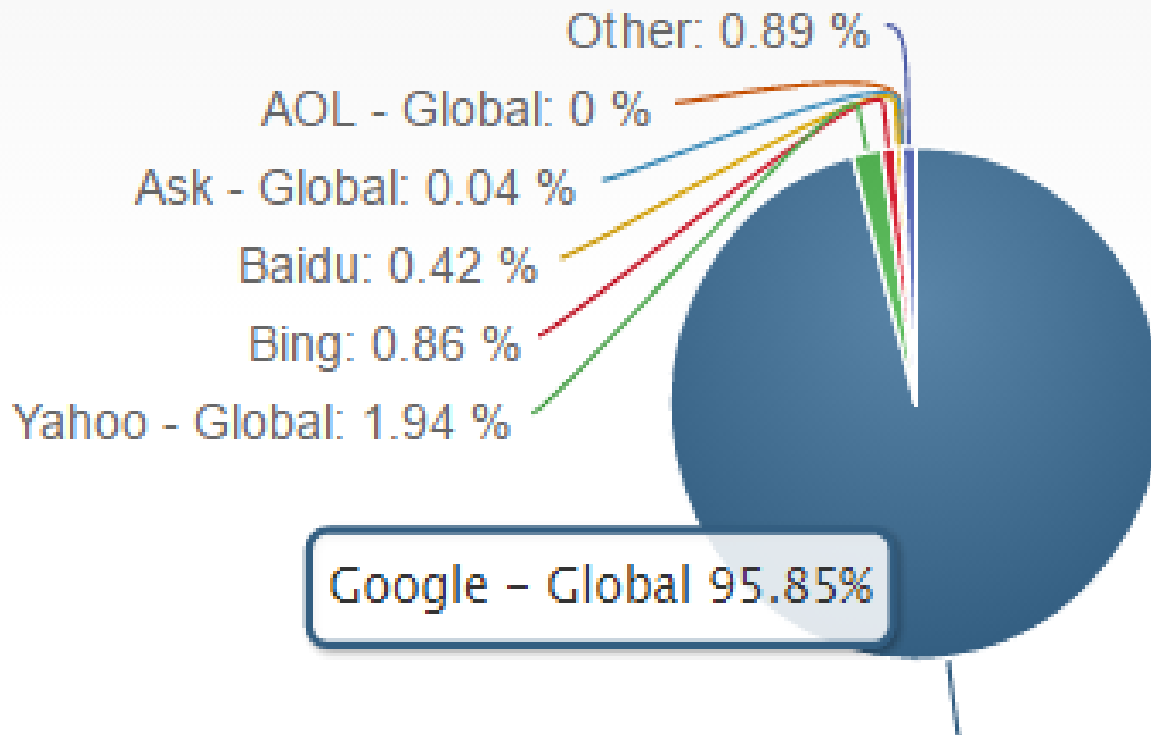
Global Marketshare –Desktop

from netmarketshare.com



<u>Google - Global</u>	80.52%
<u>Bing</u>	6.92%
<u>Baidu</u>	5.94%
<u>Yahoo - Global</u>	5.35%
<u>Ask - Global</u>	0.17%
<u>AOL - Global</u>	0.04%
<u>Excite - Global</u>	0.02%

Global Marketshare-Mobile



<u>Google - Global</u>	95.85%
<u>Yahoo - Global</u>	1.94%
<u>Bing</u>	0.86%
<u>Baidu</u>	0.42%
<u>Ask - Global</u>	0.04%
<u>AOL - Global</u>	0.00%

Webspam Glossary

- Doorway pages-Appear multiple times in results, but lead to the same destination
- Scraped content-Downloaded from another site and republished without attribution
- Auto-generated content-Written by software
- Thin content pages with affiliate links-Designed to mask scraped descriptions and/or reviews from affiliate sites.
- *Sites with any of these penalized by Google in results ranking*

Web Archiving Today

- Most archiving projects designed originally for static html text objects
- Dynamic content –Crawlers exist for this, but scalability still an issue
- Streaming Video and other multimedia- Issues of scalability, format and access
- Social Media-Most are “walled garden”, with no public access (TW is exception)
- Mobile content-Often overlooked as content is same a desktop. More AMP content will be developed that is unique to mobile.



Artificial Intelligence and Bias

*"If we want humane AI, it has to
understand all humans."*

--Megan Garcia

Not new, but growing...

- AI and machine learning enable computers to make decisions for humans
- Convenience, speed and affordability encourage more use
- Systems and algorithms **are** human creations
- Assumptions and values of those in control of AI can influence search and other outcomes, consciously or otherwise

Not new, but growing...

- Traditional secrecy around algorithms makes bias invisible
- Sources and reliability of data also subject to bias
- Machine learning is inherently self-taught and iterative
- Bias easy to perpetuate and difficult to identify and change
- “Deep learning”- Code written by machines; may not be examined by humans

Toward solutions...

<https://www.wired.com/2017/02/keep-ai-turning-racist-monster/>

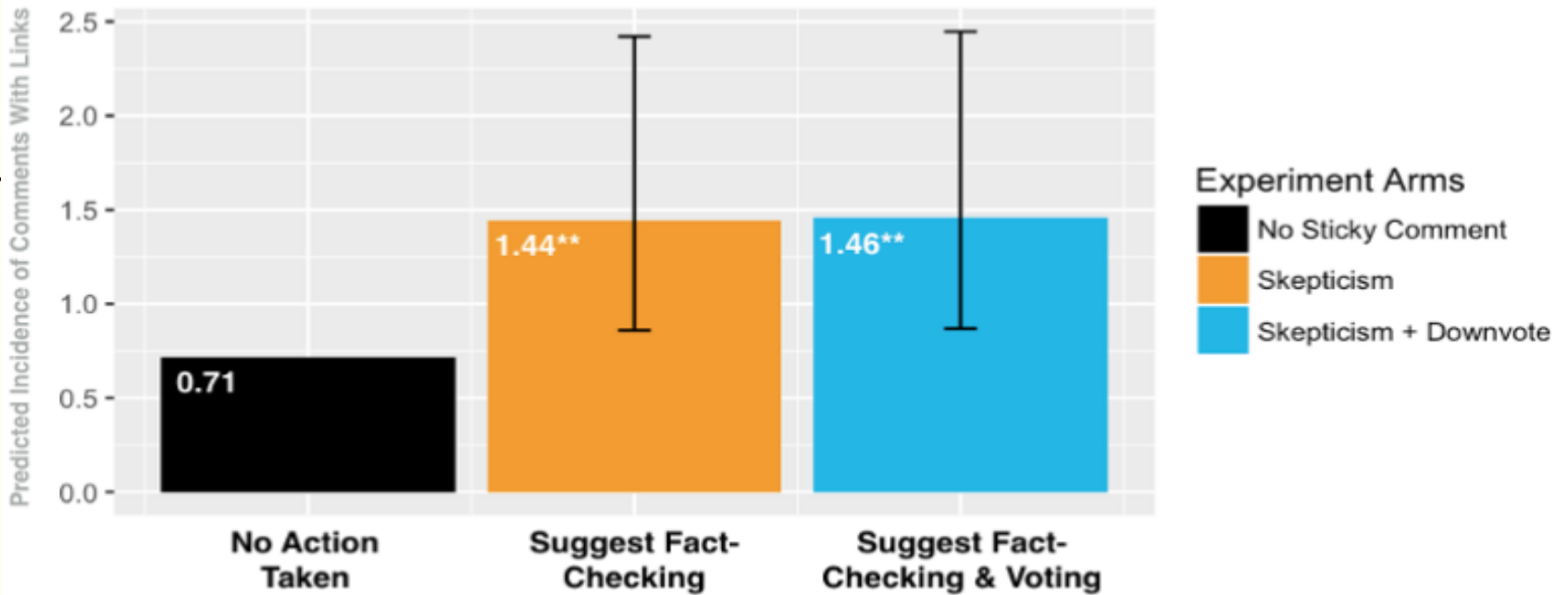
- AI development and application should include:
- Options for user communities to discover implicit/explicit AI bias and signal harassment and objectionable online behavior
- Developers from diverse groups who can spot bias before product launch
- Algorithmic auditing/testing by a third party
- Creation and support of standards to address bias across AI in research and industry

MIT's Media Lab's reddit study

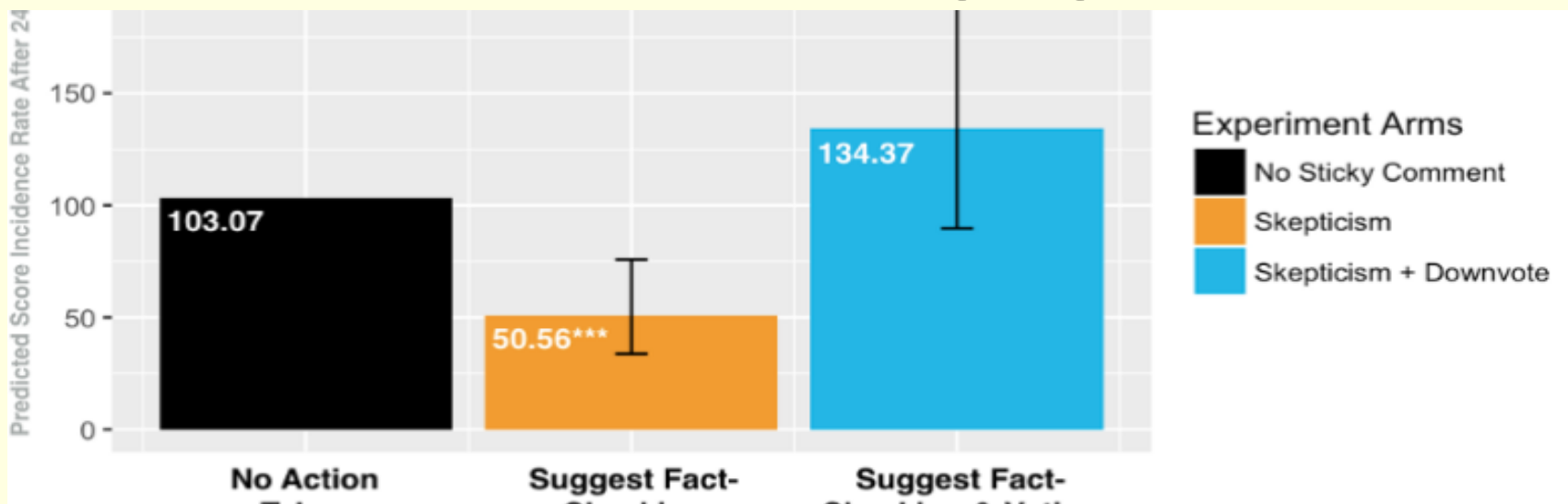
<https://medium.com/mit-media-lab/persuading-algorithms-with-an-ai-nudge-25c92293df1d>

- Reddit community was encouraged
 - 1) to fact check a set of tabloid news items
 - 2) to fact check **and** vote on another set of tabloid news items
- Study conducted 1/27/16 – 1/20/17
- AI “nudges”
 - “We can persuade algorithms to behave differently by persuading people to behave differently”

Comments with Links



Effect on reddit Ranking Algorithms



Bias in Search Engines

(adapted) <https://searchenginewatch.com/2016/02/17/search-engine-bias-an-analysis-from-the-index/>

■ Direct Manipulation

- Web crawler can ignore specific sites
- Results display can promote or penalize sites
- Political regulation and social pressures (China et. al.)
- Personalization and geo-limits of “local search”

■ Indirect Manipulation

- Algorithms based on linkage, blog activity and traffic do not account for country population differences or internet usage per capita



New Services

Digle – www.digle.com

- Hybrid people/AI answer engine
- Results from crowdsourcing, Yahoo Answers and classifieds
- Strong in commercial and local search
- Queries entered via a detailed form
- User's profile information used to match queries to "finders"
- Finders have limited time to respond
- Rewards for posting and answering determined by poster
- Requires (free) account

Peekier – peekier.com

- Privacy-based service
- Results from Bing
- No cookies or IP retention
- Search terms and history not retained
- Results displayed in preview format for greater privacy
- Tags link to related sites

Digital Commons Network

<http://network.bepress.com/>

- Showcases open access research publications
- Portal site for 275+ institutional repositories
- All content full-text and free
- Maintained by bepress, a publishing and archiving platform widely used in academia
- Subject areas:
 - Social sciences
 - Arts and humanities
 - Physical science, Math
 - Education
 - Life sciences
 - Business
 - Law
 - Medicine
 - Engineering

Splash – labs.500px.com

- Created by photography website and directory 500px
- +9 m. photos from +8 m. photographers
- Search by colors or rudimentary sketches
- Different brush sizes available
- Results clusters include landscapes, people, animals, travel and city
- Most are for sale, but results can be filtered by Creative Commons license

Start – start.csail.mit.edu

- Natural language question/answer system
- Query processing can include onscreen feedback
- Developed by MIT's InfoLab Group in 1993
- Analyzes complex queries
- Content limited

metaBUS – www.betaBUS.org

- “Cloud-based research synthesis platform”
- Free niche engine for management, organizational behavior and psychology
- +1 m. quantitative findings from reports and journals
- Arranged in +4 k topics
- Classification system facilitates correlation of variables across datasets from a range of disciplines and sources



The Social Web- Beyond the Personal

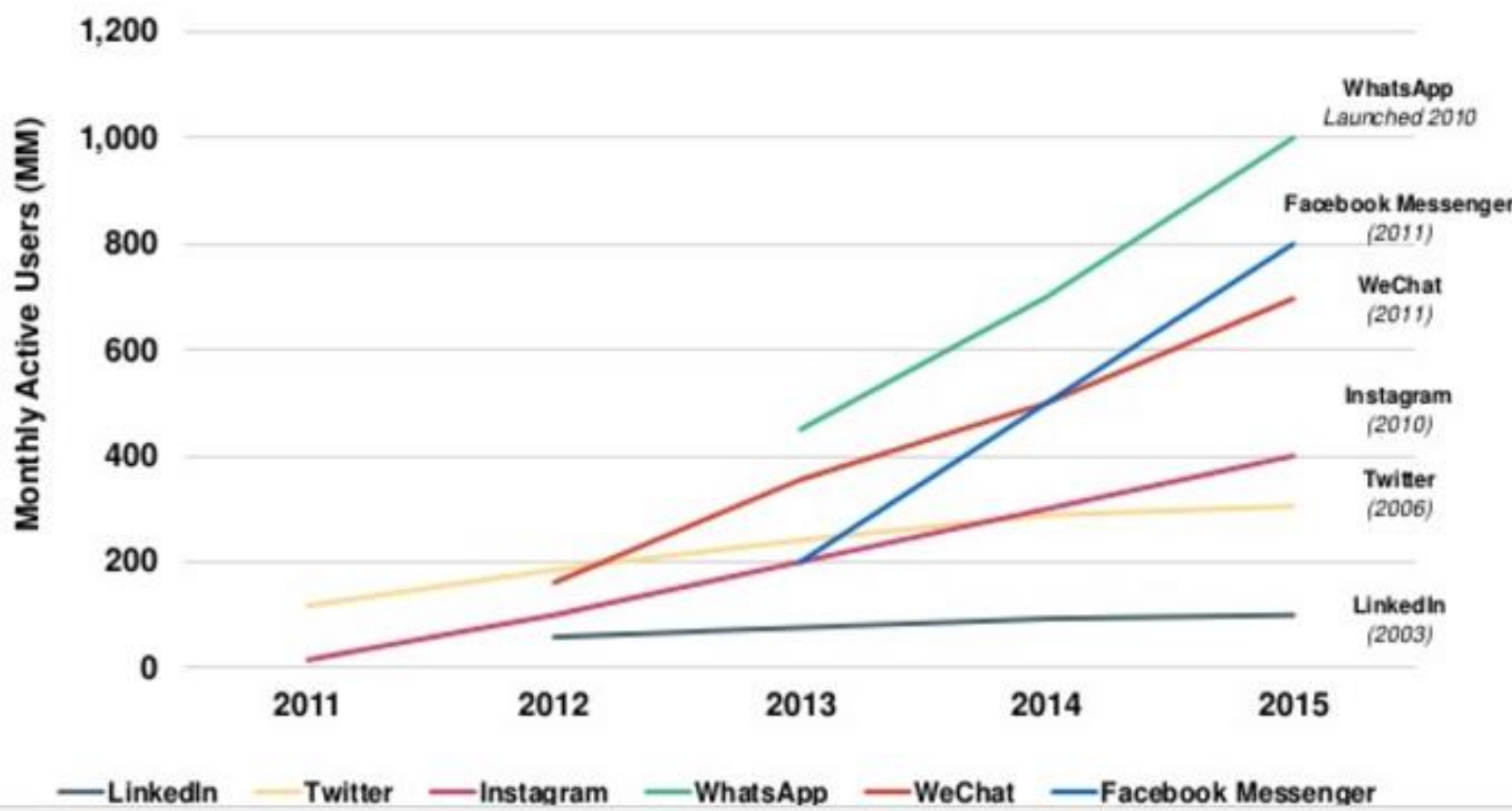
Why search the social web???

- Public responses, attitudes, opinions
 - Breaking news, events
 - Trending topics and people
 - Latest product reviews
- First-hand accounts of events-text, image, audio, video (primary sources)
- Security, technology topics (latest virus, etc.)
- Locate individuals/experts and their networks
- People interested in a topic/hobby
- Social web research projects

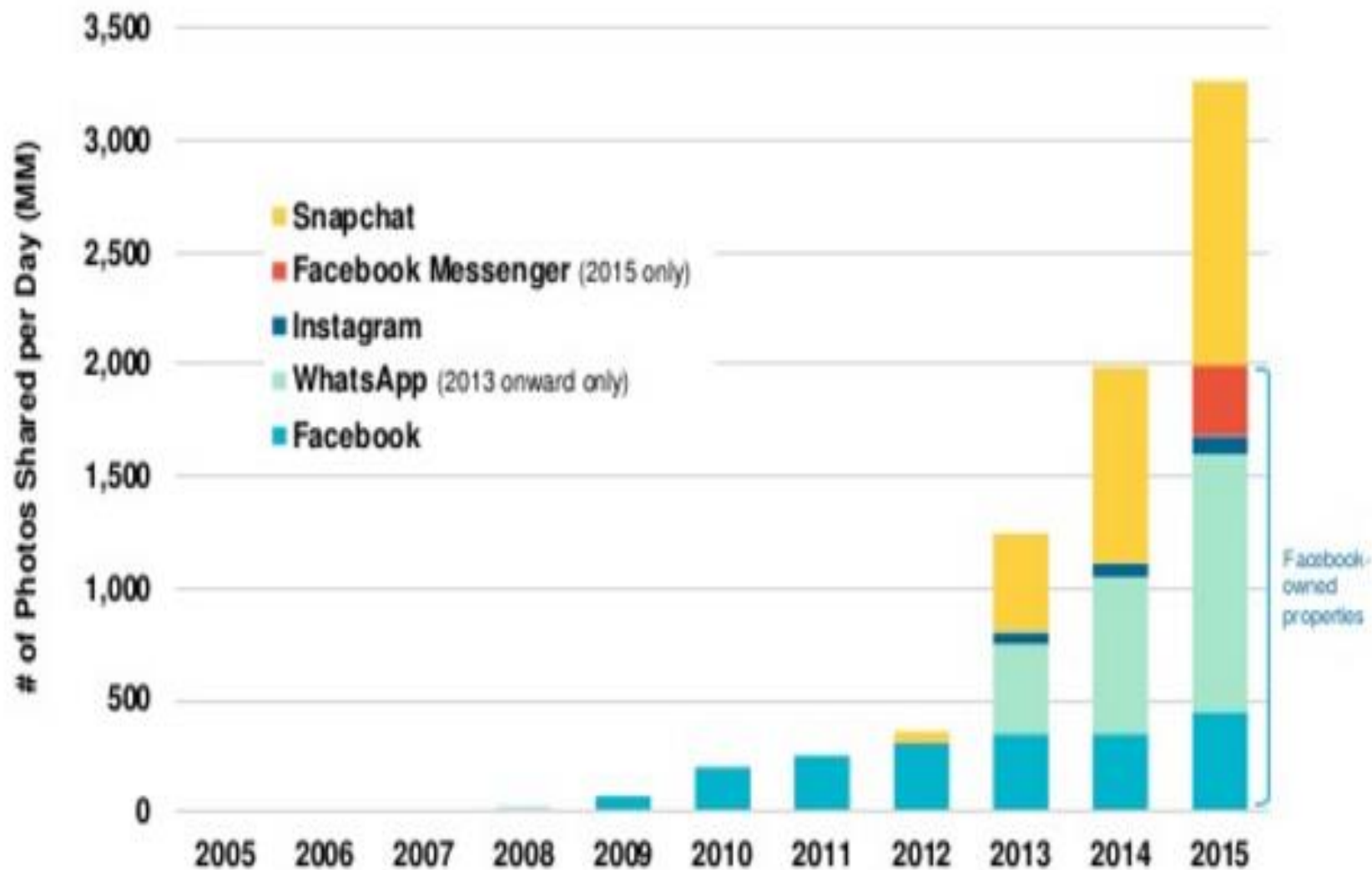
Messaging Continues to Grow Rapidly...

Leaders = WhatsApp / Facebook Messenger / WeChat

Monthly Active Users on Select Social Networks and Messengers, Global, 2011 – 2015



Daily Number of Photos Shared on Select Platforms, Global, 2005 – 2015



"Stress in America: Coping With Change"

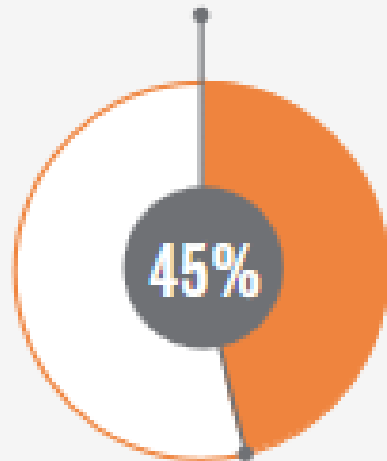
American Psychological Assn. report published 2/17

www.apa.org/news/press/releases/stress/2017/technology-social-media.PDF

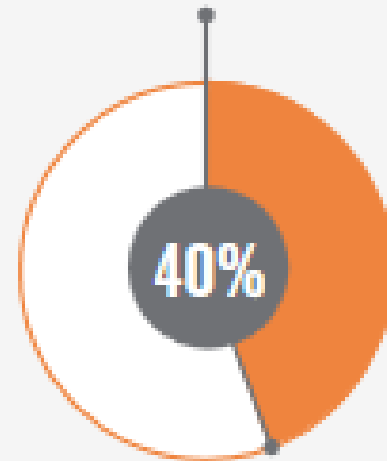
TYPICAL WORK DAY

% (of Employed) Often or Constantly Connected to at Least One Device

CONSTANTLY CONNECTED



OFTEN CONNECTED



"Stress in America: Coping With Change"

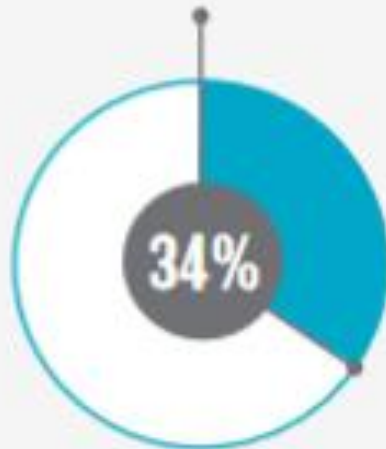
American Psychological Assn. report published 2/17

www.apa.org/news/press/releases/stress/2017/technology-social-media.PDF

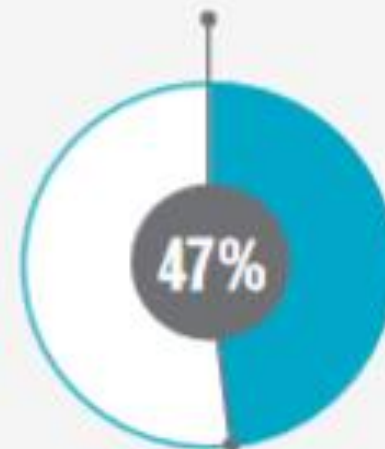
TYPICAL NON-WORK DAY

% Often or Constantly Connected to at Least One Device

CONSTANTLY CONNECTED

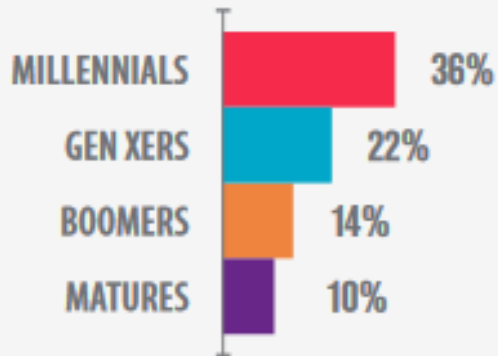


OFTEN CONNECTED

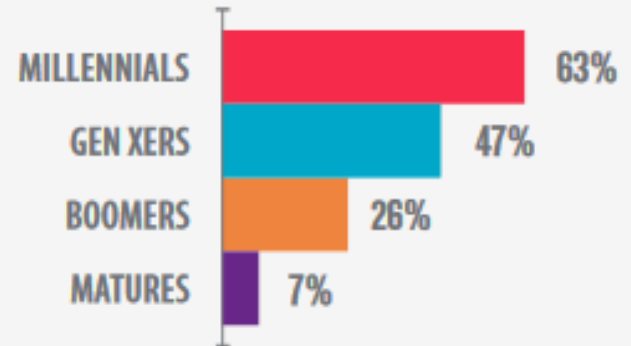




SOCIAL MEDIA HAS HELPED ME FIND MY IDENTITY

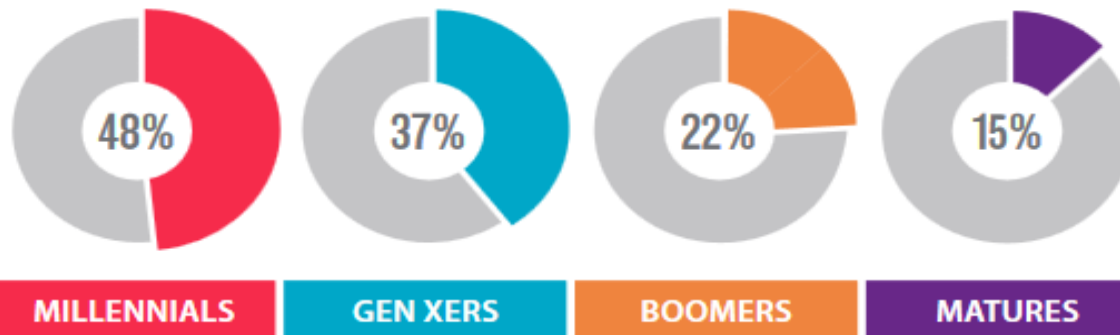


I FEEL LIKE I AM ATTACHED TO MY PHONE OR TABLET



SOCIAL MEDIA WORRIES

I worry about negative effects of social media on my physical and mental health
(% that strongly/somewhat agree)



Are we in a bubble?

“Measuring online social bubbles”

- Studies the diversity of information exposure and the impact of social networks
- “Discovery of information is being transformed from an individual to a social endeavor...”
- Bias is not *new*, but the digital era has made it more difficult to encounter ideas different from our own
- Can foster greater polarization and misinformed opinions

Are we in a bubble?

“Measuring online social bubbles”

- 3 personalized filters on Facebook
 - Our friends
 - The Feed population algorithm
 - Our own content selection
- Exposure to “the other” likely reduced by
 - 25% for conservative users
 - ca. 50% for liberal users
- Nikolov, D. et. al. (2015 December 2) Measuring online social bubbles PeerJ CompSci <https://doi.org/10.7717/peerj-cs.38>

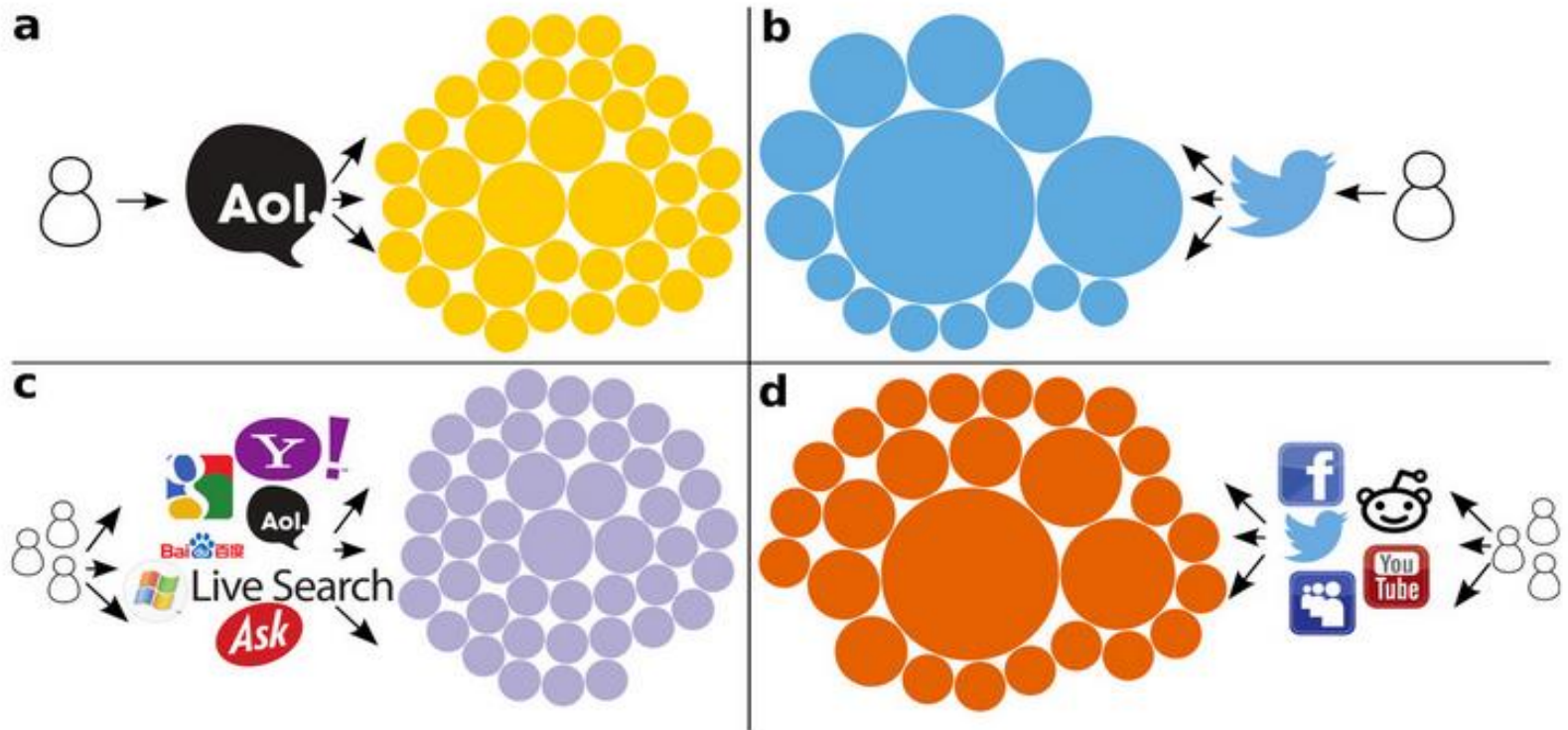


Figure 1: Diversity of information sources accessed through different online channels.

1. Each circle is a unique website; its size reflects the number of pages accessed there
2. a – Links clicked by a single search engine user
3. b – Links shared by a single Twitter user
4. c – Search traffic generated by a group of users
5. d – Social media traffic generated by a group of users

**LEFT- More heterogeneous traffic, users distributed more evenly among more sources
 RIGHT- More homogenous traffic, users distributed among fewer sources**

Facebook Search

2 trillion public posts archived - 2 billion users

- Millennials get most of their political news from FB
- Newsfeed is complex in sources and algorithms
- “How are the authors relevant to you and how are their posts relevant to your search term(s)?”
- 11 categories (in order of appearance)
 - 1) FB “Pages”
 - 2) Your friends
 - 3) News sources
 - 4) Groups
 - 5) People
 - 6) Places
 - 7) Featured posts
 - 8) Videos
 - 9) Events
 - 10) Photos
 - 11) Public Posts (incl. news sources)

Twitter Search - twitter.com/search-advanced

- Now includes every public Tweet since 2006
- Indexes ca. 3/4 trillion tweets, and grows by several billion tweets a week.
- Advanced search features
 - Booleans
 - Hashtag
 - Language limit
 - Date limit
 - Author search (tweets from or to)
 - "Near this place"
 - Attitude – positive, negative, question

Topic-based Twitter Searching

- Can provide very latest top news, tips and cutting-edge research in a topic or interest
- Increased use by world political leaders
- Locate accounts associated with the topics or issues using hashtags
 - topic or associated element
#IRS
 - person, place, event associated with topic
#worldcup
- “101 best twitter lists to follow”

<http://www.postplanner.com/101-best-twitter-lists-to-follow/>

Addictomatic – addictomatic.com

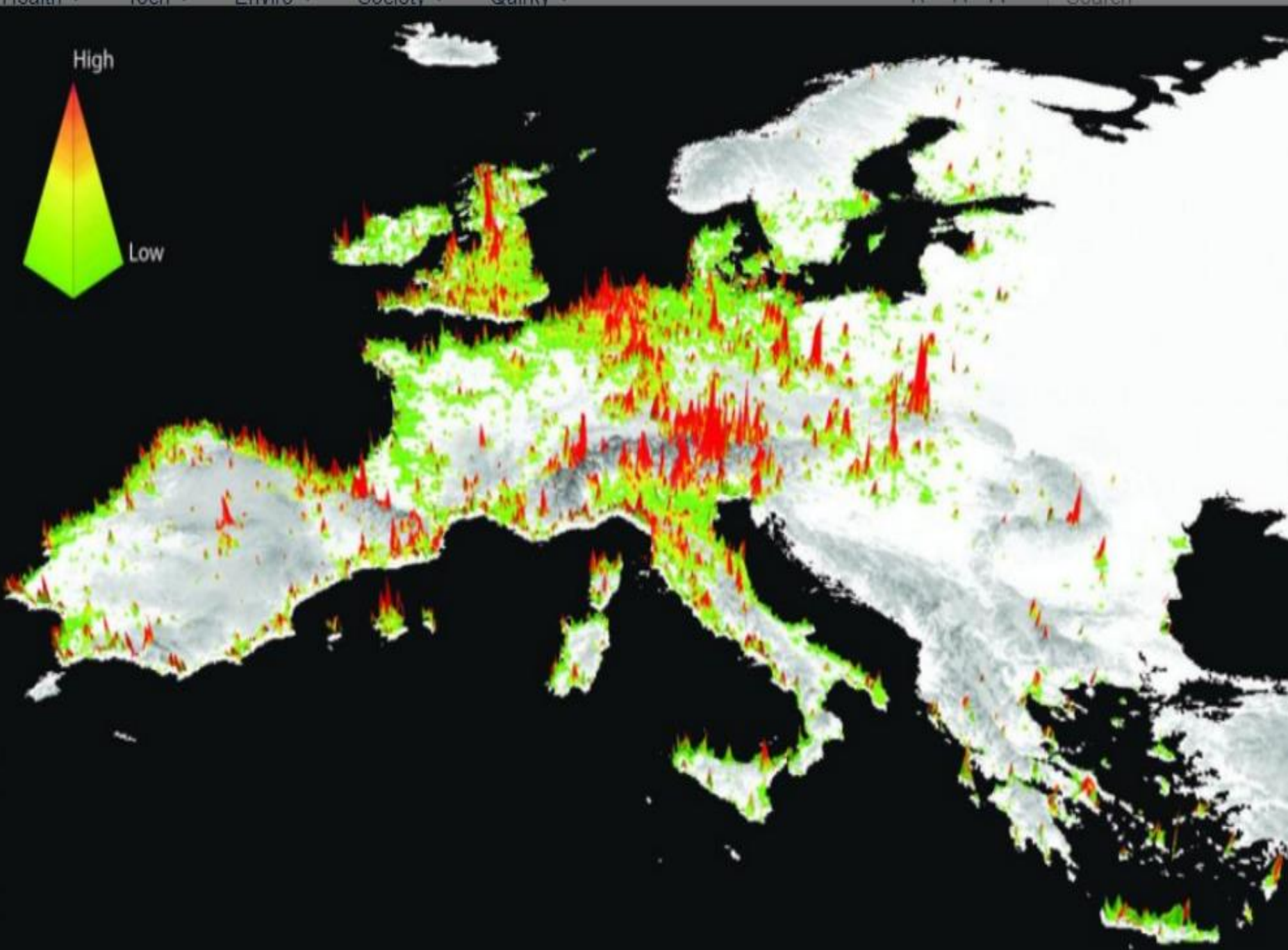
- Social search metaengine
- 25 social media and news sources
- Current topics and much more
- "Browse the news" in 24 categories
- Results vary depending on connectivity and response time(s)
- Create a custom topic page with the latest buzz on any topic covered by social media

How Are Scientists Using Social Media in the Workplace? Collins, K. et. al.

- 587 scientists from a range of academic disciplines and how they use TW
- Published 10/12/16
- Social media not yet widely adopted
- Uses (by importance)
 - Communicate with a specific colleague
 - Share their research with others in the same field
 - Stay current with breaking research in their field and related areas
 - Share research with the public and media outlets
- <http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0162680&type=printable>

Crowdsourced Predictive Models and Land Use Planning

- Set of geo-tagged photos and posts containing opinions about scenic locations in Europe
- Can inform decisions about land use
- Posts retrieved from Instagram, Flickr and Panoramio (European photo sharing platform)
- Reveal visitor sentiment, frequency and patterns
- Most valued landscapes
 - Mountainous areas
 - Rivers and lakes
 - Areas near large cities
- Continental-scale quantification of landscape values using social media data van Zenten, B. et. al. doi: 10.1073/pnas.1614158113



FB Likes as Predictors of Research Impact

Ringelhan, S. et. al. 8/5/2015

- Traditional publication review process delays assessment of a research article's impact
- Do FB likes correlate with traditional impact indicators (citations from other articles, etc.)?
- Study's sample included mss uploaded to Harvard Business School website and bioRxiv biology preprint service
- <http://dx.doi.org/10.1371/journal.pone.0134389>

FB Likes as Predictors of Research Impact

Ringelhan, S. et. al. 8/5/2015

- Correlation using total sample
 - No correlation
 - Articles with ≥ 1 FB likes have definite correlation
- Sample then divided into “psychological” and “non-psychological” content areas
- Significant positive correlation among “psychological” content
- Insignificant correlation among “non-psychological” content
- Interdisciplinary differences exist in the predictive value of FB likes on the impact of research

Social Media and Politics: Turkey 3/17

- +69,000 social media accounts identified as conducting propaganda for terrorists or insulting statesmen or the security forces following the failed coup in July 2016
- 3,894 people detained on suspicion of “making terror propaganda” on social media
 - 1,734 arrested
 - 1,328 released on probation
 - 832 remain in detention

Zimbabwe

- WhatsApp and FB used in series of protests beginning 4/16
- Citizens arrested for peaceful protests and critical media outlets shut down
- Authors of “abusive, threatening or offensive content” against the government “will be arrested”.
- Private communication under state surveillance
- Mugabe urges youth in his ruling party to use social media to “brand Zimbabwe, a Zimbabwe that is democratic, hardworking and peaceful”.

Twitter's role in modern warfare – propaganda tactics

www.bbc.com/news/technology-35842265

- Bots spam popular hashtags and use them to increase visibility of messages



Twitter's role in modern warfare –troll tactics

www.bbc.com/news/technology-35842265

- Anonymous use of TW by states or individuals- “trolling”
- Difficult to trace the origins of TW accounts
- 2/16- 2,900 TW accounts tweet a similar phrase blaming the murder of a Russian opposition leader on Ukrainians
- 2014-Russian trolls suspected of tweeting false reports of disasters in the US
- Once tweeted, false rumors are time-consuming and difficult to refute



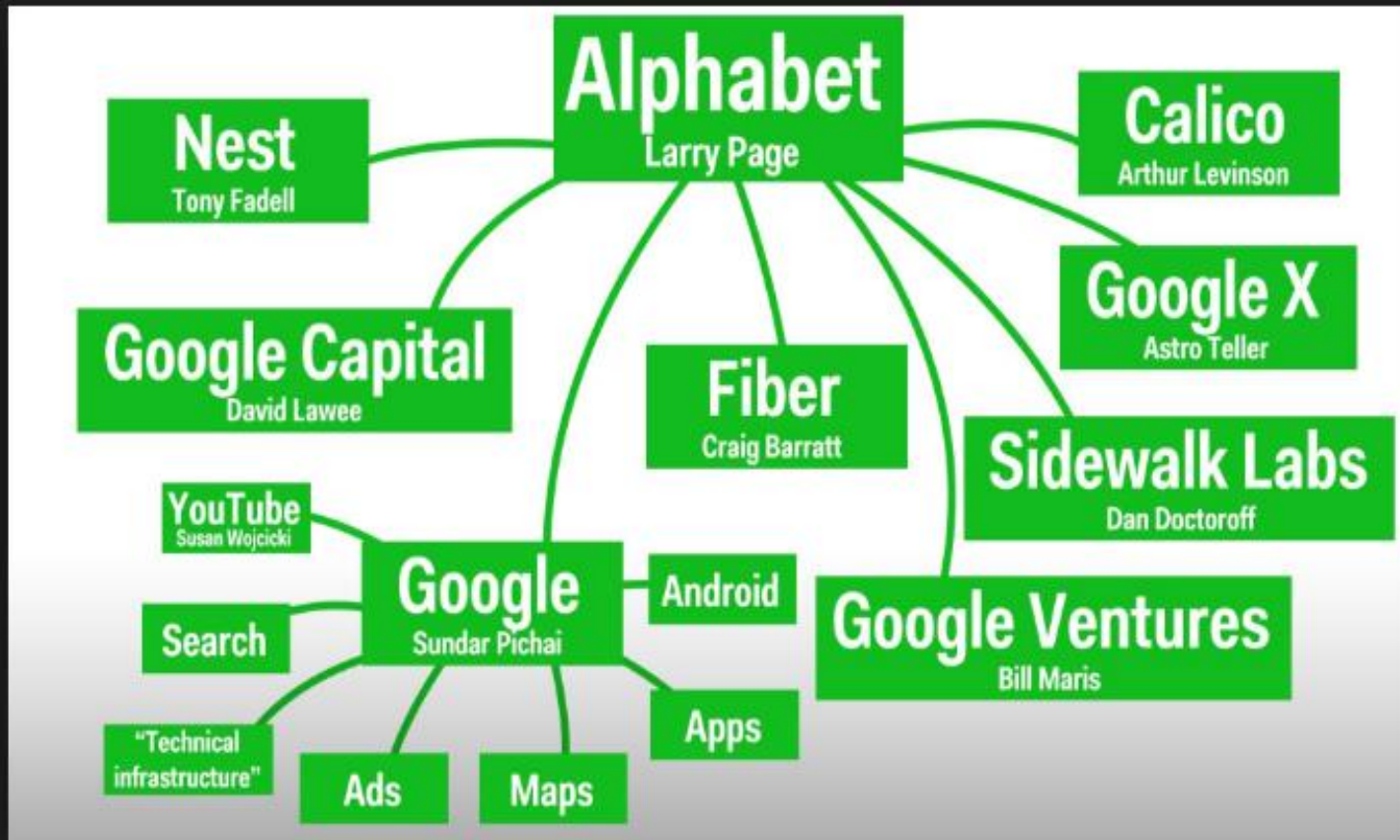
How have you made use of social media in your library

- 1) as a source of information?
- 2) as a promotional tool?



Google

Alphabet: Google Divided



Page Quality Rating

Rank rewards for content and design

- Authority
 - User-friendly site architecture
 - Contact Us page complete and always visible
 - About Us page gives purpose, history and profiles of leaders
 - Every page gives author and entity responsible
- Expertise
 - Authors qualified in the subject and credentials given
 - Sites labeled “Expert” or “Authoritative” receive more scrutiny

Page Quality Rating

- Trust
 - Original content (not scraped)
 - Amount and type of content appropriate to topic
 - Evidence of third-party endorsement
 - Contains blog that offers fresh content and forum for questions and conversation
- Linkage Affiliations
 - Incoming more important than outgoing
- Full document at
<http://static.googleusercontent.com/media/www.google.com/en//insidesearch/howsearchworks/assets/searchqualityevaluationguidelines.pdf>

Google Happenings

- 10/16 Search indexes are split between mobile and desktop
 - Mobile- mobile-friendly pages ranked higher
 - Desktop- will not be as up-to-date as mobile index
- Organic search results are now “below the fold” on mobile devices- Ads and G. cards appear first
- Rankbrain- G.’s AI system used for queries that are ambiguous or unique (ca. 450 m./day)

Google Happenings

- “Answer Boxes” now include sources recognized as authoritative in addition to Wikipedia. Most come from the top 5 organic results
- AMP (accelerated mobile pages) now in mobile News carousel- AMP is an open source format designed to load quickly on mobile devices

How To Get Started With
Accelerated Mobile
Pages (AMP)

 AMP - 1 day ago

More Google...

- Google expands hardware products with Pixel phones, VR Viewer, Home and more
- Home- Voice-activated control network for media, lights, thermostats and switches (via G Assistant)
- G Earth VR upgrade allows for 360° views for certain locations
- G's 2017 Oscar nomination for its 6-minute animated Pearl. The work is an AI created VR film about a father-daughter road trip.



Latest From Other Established Services

DuckDuckGo – ddg.gg

- New search box commands
 - ! *Search only the specified resources*
 - !w *Search only in Wikipedia*
 - !google *Google results free of personalized filters (they don't know it's you)*
- Other special commands
 - **Is (a website) down** - site status
 - **qr** - Creates a QR code for the site
 - **expand (shortened url)** - does so
 - **password (length) strong** - creates it
 - **color codes** - codes for onscreen colors

Bing's Conversational Search

- Connects subsequent query to your first one if they can be related
- Functional for 1 subsequent query only
- G processes each query as independent of another
- Privacy controls now at bottom of results page and much more accessible

Startpage – www.startpage.com

- Privacy engine now merged with Ixquick metaengine
- No tracking or profiling of searches
- Results from Google but searches not tracked by Google
- European based
- Shielded from US data collection mandates
- Search terms and data not shared with any other service
- Ixquick (www.ixquick.eu) still running

Future of Search

My predictions last year...

- Impact of mobile devices will drive R&D and influence developments in desktop search
- Social networks and crowdsourced search engines will grow in quality, popularity and political influence
- Content from apps will become more accessible, from inside and outside the app
- Predictive search and personal virtual assistants will grow
- Governments will continue to challenge search engines in areas of monopoly and individual privacy

Future of search

This year's predictions

- Personal assistants and voice search will grow in search and daily life
- AI will come closer to understanding meaning, intent and emotion in text, images and video
- Mobile search considerations will lead developments in web search engines
- Paid and engine-sponsored results such as G's Cards will appear over organic results
- The Social Web will become more content-significant and vital to human communication of all types

*Thank You and
Enjoy Your Searching!*

Michael Hunter

Reference Librarian (part time)

Hobart and William Smith Colleges

Geneva, NY 14456

(585) 770 8451

hunter@hws.edu