Big Data and Social Media on the Internet

Vint Cerf
CERN
Geneva

June 2018
The Original ARPANET
Dec 1969
First Three-Network Test of Internet

November 22, 1977
Social Media

• Early evidence of social effects: ARPANET mailing lists
  • Sci-Fi Lovers, Yum-Yum

• Engelbart’s oNLine System (NLS)
  • Shared documents, hyperlinking, mouse, portrait mode displays, collaboration

• USENET (Tom Truscott and Jim Ellis conceived the idea in 1979, and it was established in 1980.)

• First spam!
  • Gary Thuerk, DEC, job advertisement – ARPA/IPTO nuclear reaction

• The WELL was started by Stewart Brand and Larry Brilliant in 1985, and the name (an acronym for Whole Earth 'Lectronic Link)

• PRODIGY, AOL. COMPUSERVE, hundreds of Bulletin Boards
  • Most reached by dial-up modems in the 1980s
Big DATA

- Astronomical databases and imagery
  - Hubble, Radio telescopes, Large Synoptic Survey Telescope,…

- Collider data (e.g. LHC!!!)
  - Massive, and rapid filtering required

- Genetic information (Human Genome, now others)
  - Proteome, Microbiome, Metabolome

- Pre-WWW (Wide Area Information System, Gopher, Archie…)

- WWW indexing (AltaVista, Yahoo!, Google, Bing, Baidu, …)

- Twitter, Instagram, FaceBook, Google+, Snapchat, WeChat, QQ…

- E-Commerce data (Amazon, eBay, YELP, Alibaba, …)

- Environmental data of all kinds (atmospheric, seismic, water, …)
Taming Big Data

- **Google: Big Table, Spanner**
  - Requiring massive computing data centers

- **Multi-layer Neural Networks**
  - Requiring superfast training and processing (e.g. TensorFlow)
  - Brittle and potentially chaotic results (adversarial images)
  - Dramatic results (AlphaGo, AlphaGo Zero, Data Center cooling, self-driving cars, natural language translation - German weather story)

- **Internet of Things data avalanche**
  - Safety, Security, Privacy, Reliability, Autonomy
  - Examples of IOT side effects

- **Dynamic, multi-variate presentations** (e.g. Gapminder)
Example: Gapminder multivariate presentation
New Challenges

• Misinformation, Disinformation, Critical Thinking

• Positive and negative social media feedback loops
  – Bullying, stalking, doxing, addiction to feedback
  – Rewarding socially beneficial activism, corporate social responsibility

• Digital Literacy

• Digital Preservation

• (Inter)national Law Enforcement/Processes

• Consequences of Malware and Buggy Software

• Ethics and Software-based Systems
Unfinished Business

• Technology development
  • Reliable and updateable software
  • Strong, End-End Authentication
  • Confidentiality and Privacy Protection
  • IPv6 Address Implementation
    • Internet of Things, Cyber-Physical Systems, Mobiles
  • Long-Term Digital Preservation
  • Stable Identifier Systems (Beyond Domain Names)
  • Broadband Wired and Wireless Access
  • Open Software Defined Networks (switch + software)