## Closing remarks

1

#### Where we started

"Google's mission is to organize the world's information and make it universally accessible and useful" Google's mission statement, ~ 1998.

World Wide Web invented by Tim Berners-Lee 1989

"A memex is a device in which an an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory." Vannevar Bush, As we may think, *Atlantic Monthly*, July 1945.

2

## Where we have been: major themes

3

#### 1. Mathematical models fundamental

- how model information
  - Capture structure within?
- vector models + linear algebra
- graph models
  - · links and paths
- probabilistic models
  - Markov models
- · applications: text, Web, other media

## 2. Algorithms & data structures

- indexing
- graph analyses
- random walks
- eigenvector computation
- clustering
- sampling
- aggregating
- compressing

#### Algorithms meet BIG practice

distributed computation

#### 4. Evaluation

- Must have quantified
  - Not just "see how well works"
- · May need more than one measure
- Quantifying does not preclude human perception

3. Costs: what optimizing?

- Quality of solutions
  - Almost always approx' tions or heuristics
  - Humans add information
    - · users: characteristics & feedback
    - · authors: semi-structured content
- Performance
  - Disk I/O
    - · Drives times of algorithms
    - Drives main memory needs buffers, caches
  - Network latency
    - highest time and \$ cost

6

What did we miss?

# Information Summarization and Visualization

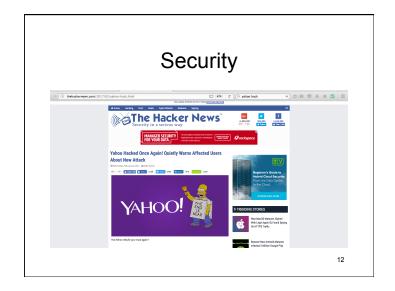
For illustrations of Metro Maps, see <u>Trains of Thought: Generating Information Maps</u>, Dafna Shaha, Carlos Guestrin, and Eric Horvitz, *Intern. World Wide Web Conf. (WWW)*, ACM, 2012, pp. 899-908.

9

For illustration of social graph of feelings see

We Feel Fine and Searching the Emotional Web, Sepandar D. Kamvar and Jonathan Harris, *Proc. of the Intern. Conf. on Web Search and Data Mining* (WSDM), ACM, 2011, pp. 117-126.





## Where are we going

13

## 2. search everything

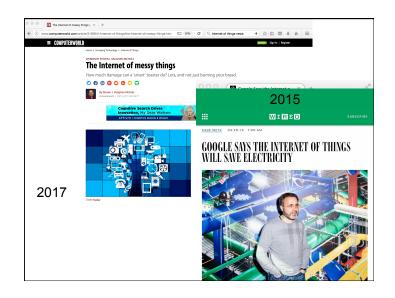
- multi-media
  - obvious: image, music, video,
  - maps
  - other?
- social networks
  - Influence in many ways
- cloud computing
- objects: "internet of things"

15

### 1. new modes of search

- · real-time search
- non-text based input
  - spoken requests
  - content-based input
- concise output
  - text versus non-text
  - summarizing
    - single source
    - multiples sources

handhelds drive much of



#### 3. real semantic-based search

" ... It is an enlarged *intimate* supplement to his memory."

Vannevar Bush

- · question answering
- · understanding user intent

17

## Where are "we" going?

 $1 + 2 + 3 + ? \rightarrow Semantic Web ?$ 

vision of use of information by active agents
Berners-Lee, Hendler and Lassila
Scientific America, May 2001

18

#### Semantic Web Overview

- Initiative of W3C: WWW Consortium
  - academic, government and industry
  - begun 1994 by Tim Berners-Lee
- · common frameworks for data specification
- · frameworks allow sophisticated functionality
  - automated understanding and use of information
  - Inference!
- · open specifications, open source
  - Allow independently written tools interoperate

9

## Major concerns going forward

- Security!
- Privacy!
- · Data explosion?
- Universal access?
  - Resource limitations
    - developing nations
- Others?

 For the statistics on world internet usage and percent penetration of the population by region see <u>Internet Usage Statistics</u>, *Internet World* Stats