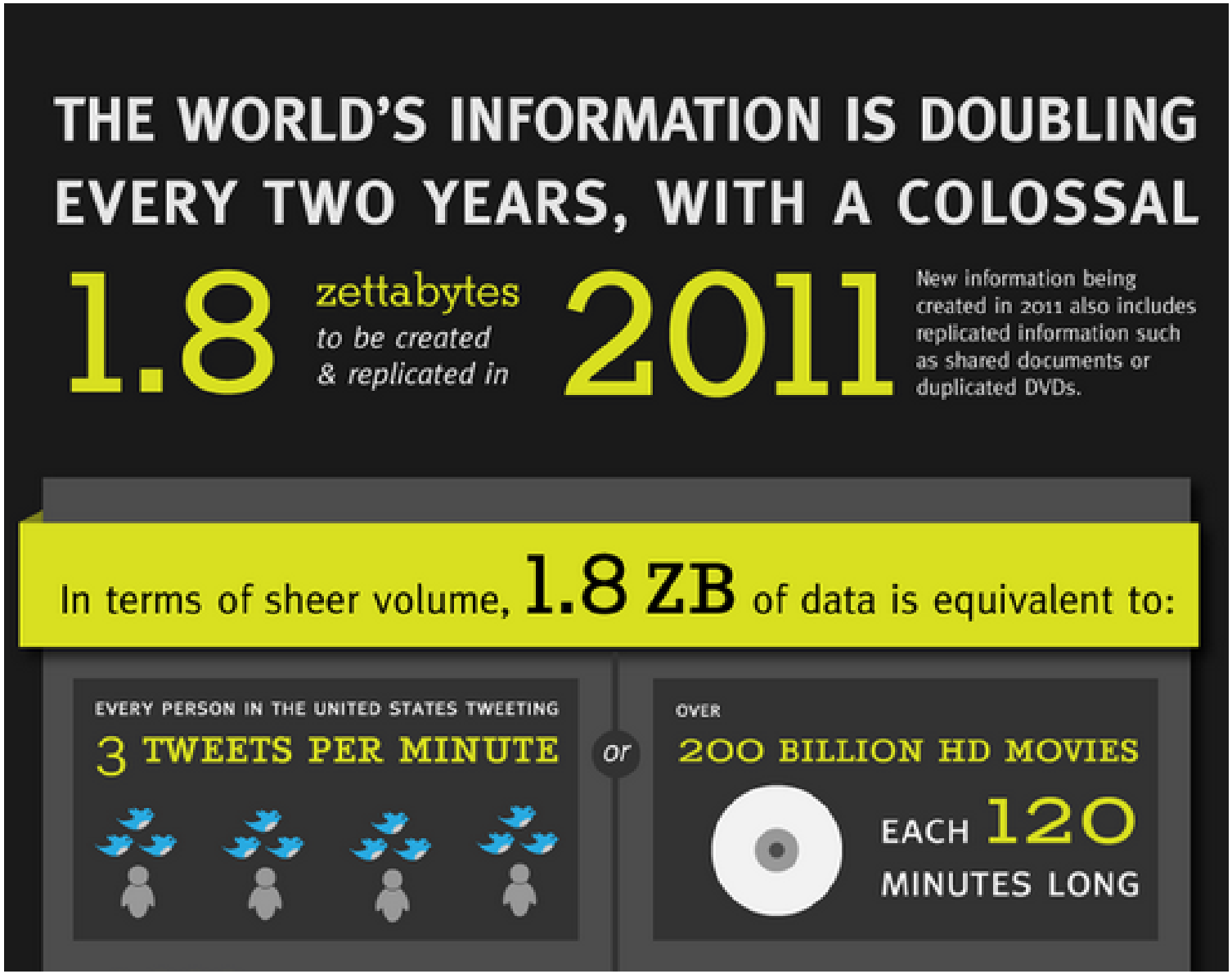
What about big data?

Jeffrey Leek

May 17, 2016

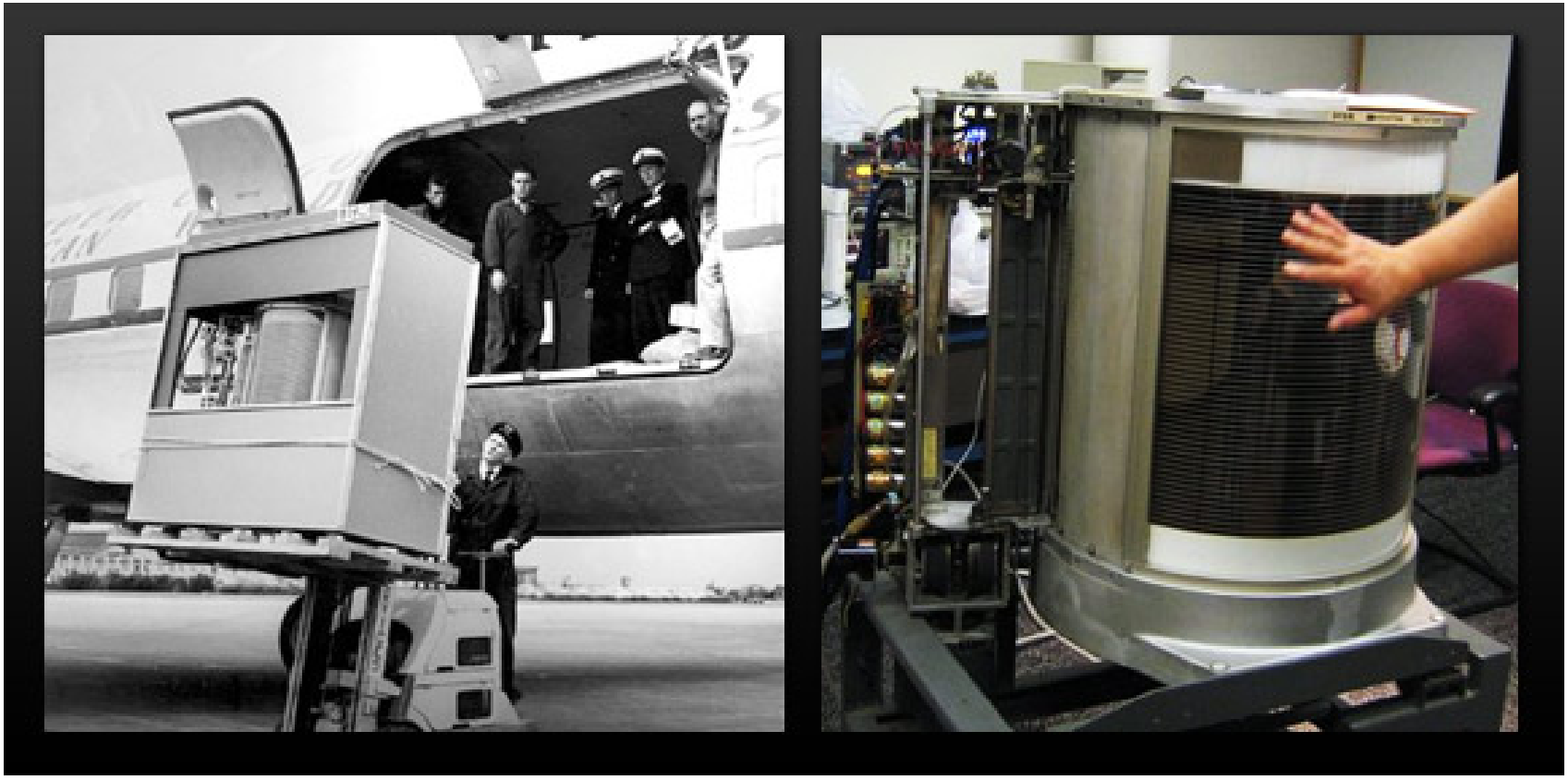
How much is there?



http://mashable.com/2011/06/28/data-infographic/

So what about big data?

Depends on your perspective



Why big data now?

An Experimental Study of the Small World Problem*

JEFFREY TRAVERS

Harvard University

AND

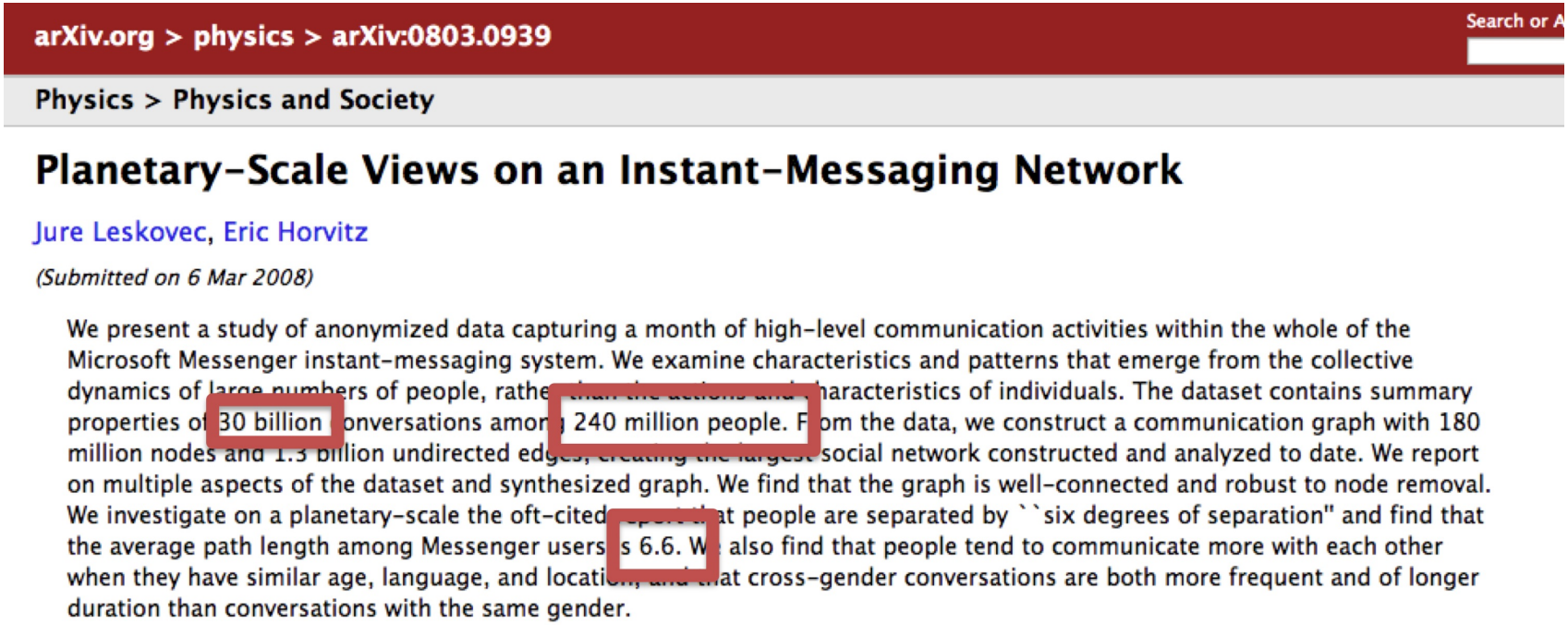
STANLEY MILGRAM

The City University of New York

Arbitrarily selected individuals (N=296) in Nebraska and Boston are asked to generate acquaintance chains to a target person in Massachusetts, employing "the small world method" (Milgram, 1967). Sixty-four chains reach the target person. Within this group the mean number of intermediaries between starters and targets is 5.2. Boston starting chains reach the target

Travers and Milgram (1969) Sociometry

Why big data now?



arXiv.org > physics > arXiv:0803.0939

Search or A

Physics > Physics and Society

Planetary-Scale Views on an Instant-Messaging Network

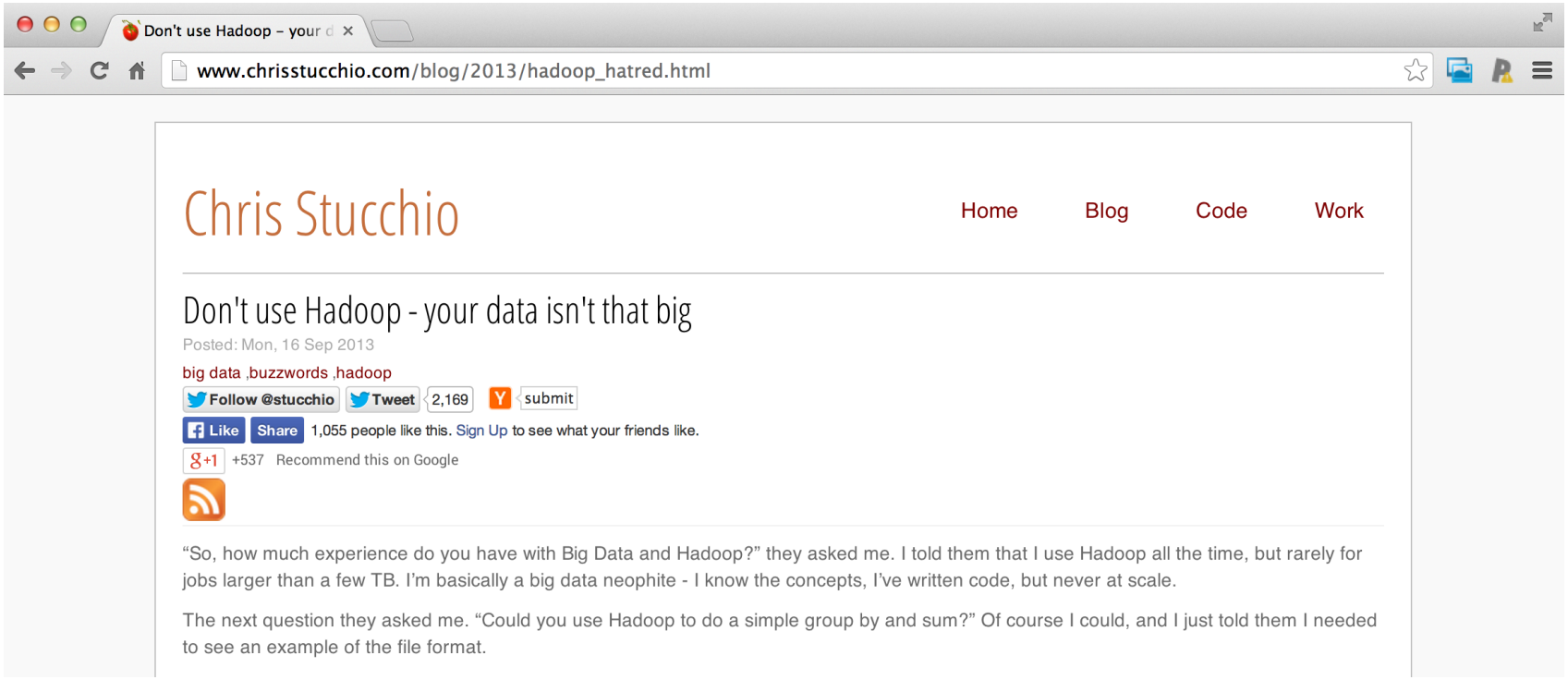
Jure Leskovec, Eric Horvitz

(Submitted on 6 Mar 2008)

We present a study of anonymized data capturing a month of high-level communication activities within the whole of the Microsoft Messenger instant-messaging system. We examine characteristics and patterns that emerge from the collective dynamics of large numbers of people, rathe haracteristics of individuals. The dataset contains summary properties of 30 billion conversations amor 240 million people. F om the data, we construct a communication graph with 180 million nodes and 1.3 billion undirected edg social network constructed and analyzed to date. We report on multiple aspects of the dataset and synthesized graph. We find that the graph is well-connected and robust to node removal. We investigate on a planetary-scale the oft-cited at people are separated by ``six degrees of separation'' and find that the average path length among Messenger users s 6.6. W also find that people tend to communicate more with each other when they have similar age, language, and locati at cross-gender conversations are both more frequent and of longer duration than conversations with the same gender.

Leskovec and Horvitz WWW '08

Big or small - you need the right data



`http://www.chrisstucchio.com/blog/2013/hadoop_hatred.html`

Big or small - you need the right data

The data may not contain the answer. The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data. . .

Tukey

Big or small - you need the right data

. . . no matter how big the data are.

Leek