Social Networking-based Systems

Stefan Saroiu and Nick Koudas

Dept. of Computer Science
University of Toronto

Social Nets. Systems@U. of T.

- **Mission**: Build large-scale Internet systems that enable mass communication leveraging social networks

- High-level research themes:
  - Access control
  - Search
  - Content delivery and caching
- Young and energetic research group:
  - 4 faculty + ~15 graduate students

This Talk

■ Two systems:

1. A social networking-based access control scheme for personal content

2. Social media aggregation

This Talk

- Two systems:

1. A social networking-based access control scheme for personal content

2. Social media aggregation

A Social Networking-Based Access Control Scheme for Personal Content

Stefan Saroiu
U. of Toronto

Sharing Personal Content Online is a Mess!

- Huge growth of personal content
  - Photos, videos, blog pages, song collections

- Two common ways of sharing it with your friends and family

1. Use e-mail service (e.g., GMail)

- Restrictions due to GMail's bandwidth costs
- Email is based on Push model
  - Inefficient way to share content

- E-mail was never designed to act as a content delivery system!

2. Use Social Networking Sites

- Many content-specific sites with little support for access control

- Users must maintain many copies of social network
    - Copies of social network become inconsistent and out-of-date
    - Users must be registered to upload/access restricted content
    - Redundant information in many cases

- **People have one social network, yet they now must manage several copies of their social network**

Sharing Personal Content Online is a Mess!

- There is a mismatch between people's needs and what today's sites offer!

Separate Social Information from Content Serving

- Observation: today's sites store **both** social information **and** personal content
- Insight: separate them
    - Let people manage social networks
    - Let sites serve content

Our Proposal
for Cleaning Up the Mess

Let People Manage Social Networks

- Capture social relations digitally with social attestations
    - Issue signed attestation to your friends, family, buddies, etc…
    - People use attestations to prove their relationships to others

- Maintain **one** social network – personal address book
    - Managed by user
    - No infrastructure component

Let Sites Serve Content

- Sites can offer access control suitable to the needs of personal content
  - Personal content needs social networks

- Social ACLs:
  - Restrict access to private photos to people with "family" attestations
  - Restrict access to private song collections to people with "friend" attestations

New Internet Applications

- Social BitTorrent
  - Share personal content over BitTorrent
  - Restrict access to content to friends only
- Social firewalls
  - Restrict home machine access to friends only
- Social Google Calendar
  - Publish different views of online calendar
    - One view to family, another view to colleagues

Current Status of Our Systems

- Adding our access scheme to Flickr
  - Working implementation
  - Upcoming initial release: November 2007
- Social BitTorrent
  - Upcoming initial release: December 2007
- Social Google Calendar
  - Upcoming initial release: January 2008

This Talk

■ Two systems:

1. A social networking-based access control scheme for personal content

2. Social media aggregation

This Talk

- Two systems:

1. A social networking-based access control scheme for personal content
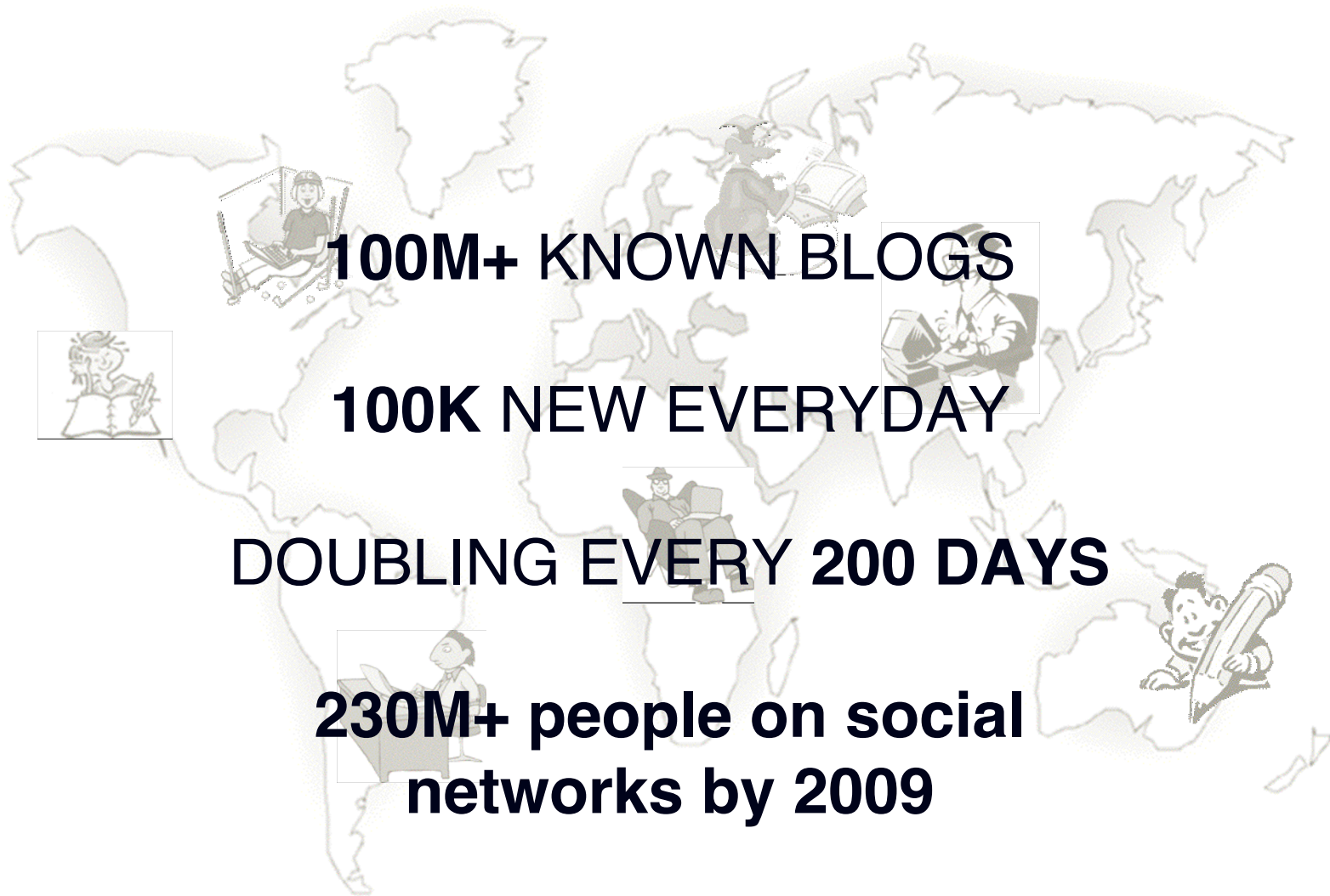

2. Social media aggregation

Social Media Aggregation

Nick Koudas
University of Toronto

BLOGOSPHERE

100M+ KNOWN BLOGS

100K NEW EVERYDAY

DOUBLING EVERY **200 DAYS**

230M+ people on social networks by 2009

WHAT ARE THEY WRITING ABOUT??

PERSONAL LIFE
PRODUCT REVIEWS
POLITICS
TECHNOLOGY
TOURISM
SPORTS
ENTERTAINMENT

WHY SHOULD WE CARE?

HUGE DATA REPOSITORY

WILL CONTINUE TO **GROW**

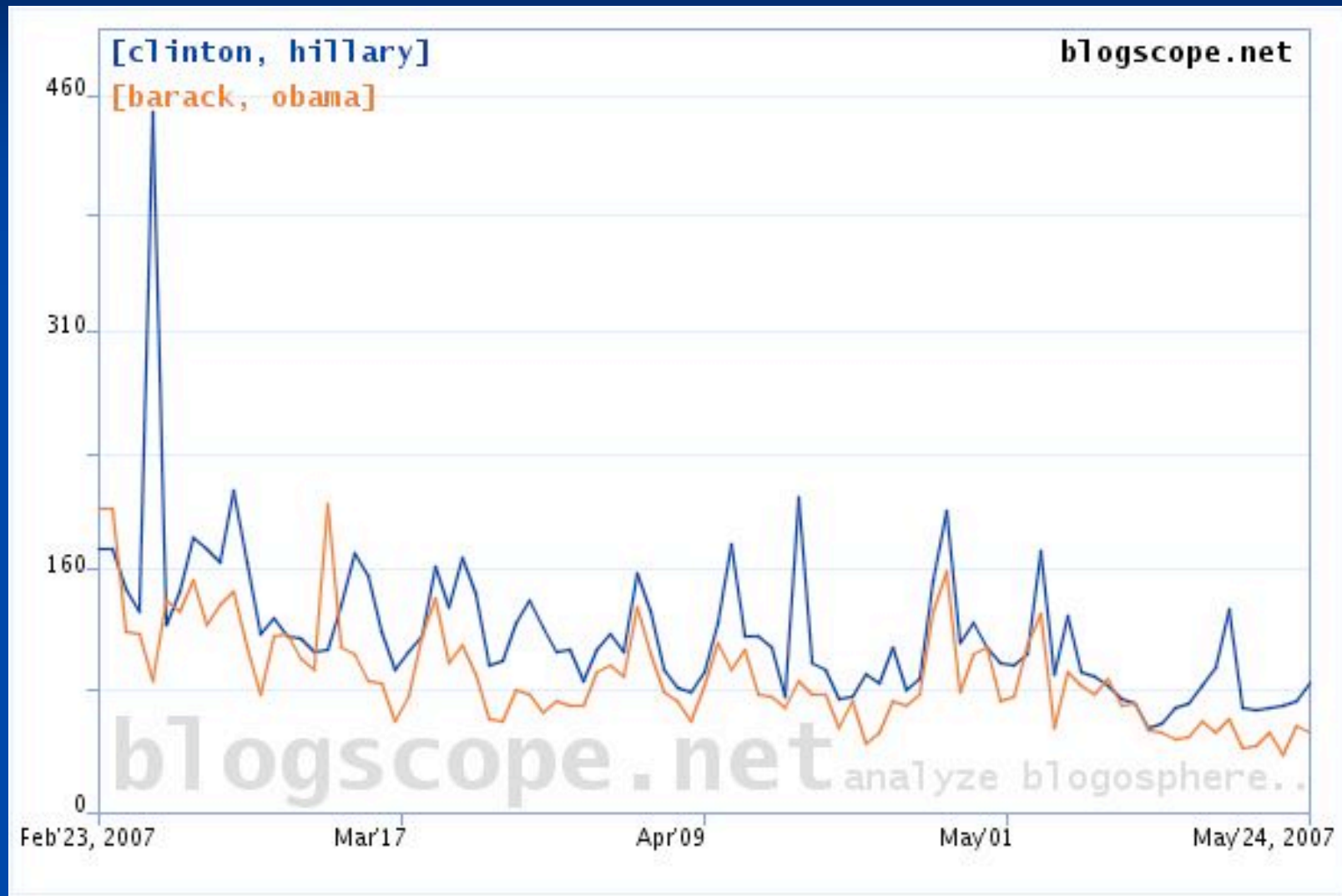EXTRACT **PUBLIC OPINION**
(what are my customers talking about?)

VALUABLE INSIGHTS

CHALLENGES AND OPPORTUNITIES

HUGE AMOUNTS OF **UNSTRUCTURED** TEXT
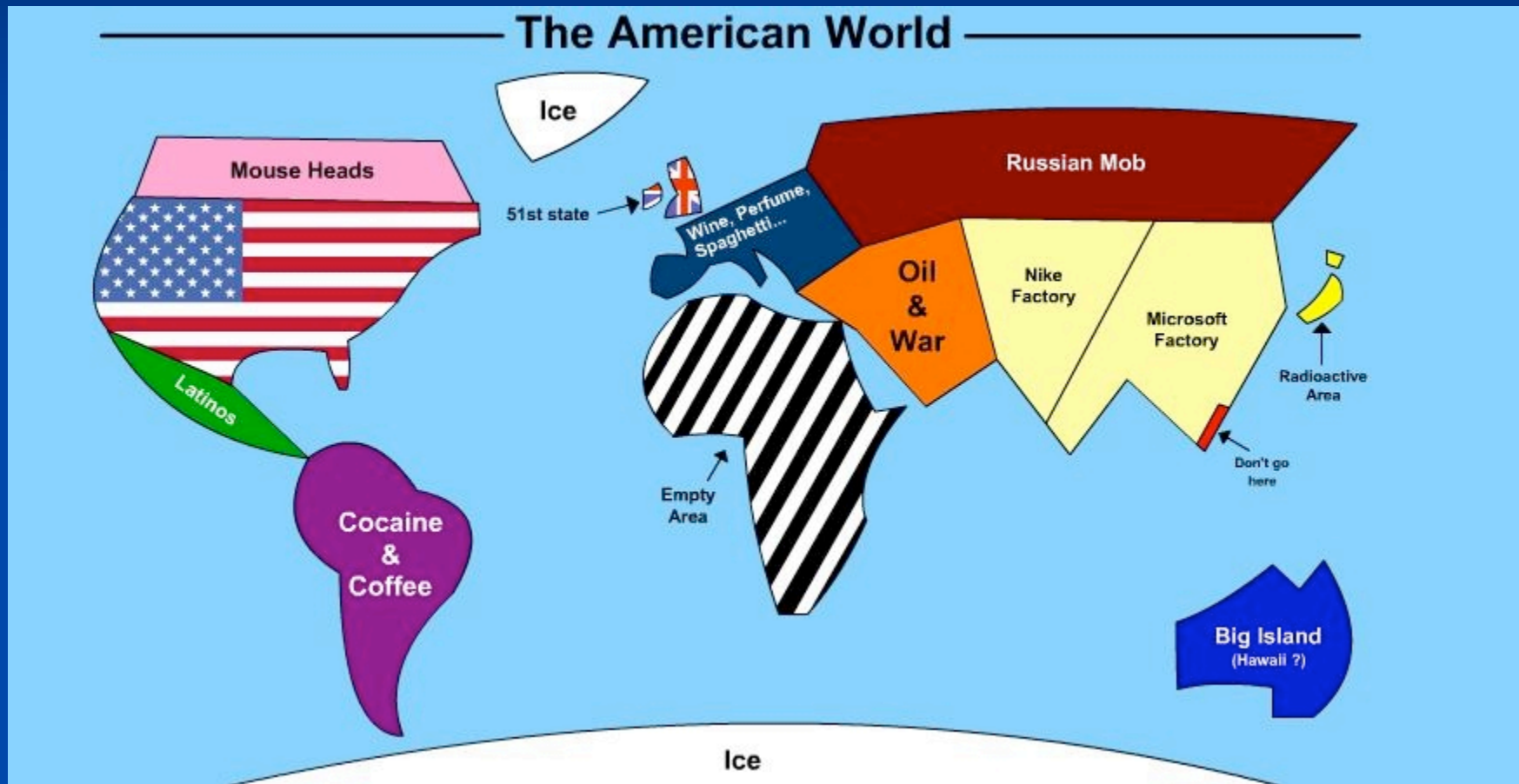
TEMPORAL DIMENSION

GEOGRAPHICAL ASSOCIATION

CONVERSATION

BLOGSCOPE

CRAWLER RUNNING 24x7

TRACKING **14M** BLOGS
(will cover all blogs by year end)

INDEXING **120M** ARTICLES
(will increase by an order of magnitude)

AGGREGATION AND PREPROCESSING

INTERACTIVE SEARCH AND **ANALYSIS**

ANY **STREAMING TEXT SOURCE**

NEWS

MAILING LISTS

FORUMS

SOCIAL MEDIA

SOCIAL NETWORKS

Opportunities

- Community extraction (based on content)
- Conversation topics, chatter
- Distill opinions, semantic orientation
- Blogger (customer) segmentation based on profile information (age, gender, etc)
- VISIT: www.blogscope.net

Shameful Plug

- Come join our group if you want to work on building these cool systems

More Info

- http://www.cs.toronto.edu/~stefan
- stefan@cs.toronto.edu

- http://www.cs.toronto.edu/~koudas
- koudas@cs.toronto.edu