Engineering a Common Good

Fair Use of Aggregated, Anonymized Behavioral Data

Nathan Eagle
The Santa Fe Institute
nathan@santafe.edu
Part I
Client-Side Data Collection: Individual and Group Behavior
N = 1 – 10,000
Sex Workers in Mtwapapa

• 12 Nokia N70s with Context logging software
  - Location, Bluetooth, Communication
• Question: Can we identify the onset of a high-risk situation?
• Can we use mobiles as a tool for education and the dissemination of context-specific information?
Children in Kilifi

• Respiratory Syncytial Virus (RSV)
  • Airborne pathogen prevalent among children under 5
• 50 Households – Phones and RFIDs
  • Location, Bluetooth, Communication, Close Proximity (RFIDs)
• Question: Can we link proximity patterns with infection?
Medical Field Workers in Kilifi

• 30 Medical Field Workers with mDSS Application
  • Location, Real-time Database Questionnaires
• Question: Can we replace paper-based surveys with mobile phones?

Smokers in New York City

- 150 Undergraduate Smokers / Recent Quitters
  - Location, Bluetooth, Communication, Context-Driven Surveys
- Question: Is there a behavioral signature associated with relapse?
Part II
Server-Side Data Collection:
Cities & Nations
N = 10,000 – 300 Million
Server-Side Data Collection: Call Data Records (CDR)

- **CDR Column Elements**
  - Caller (Texter) / Receiver
    - Hashed ID!
  - Time / Duration
  - Cellular Towers

- **Typical Data**
  - 1 Month – 5 Years
  - 5% – 100% of the population
Server-Side Data Collection:
Call Data Records (CDR)

- Additional Edges
  - Me2U Airtime Transfers
  - ‘Please Call Me’ free texts
  - Financial Transfers (mPesa)

- Individual Attributes
  - Product / Tariff Adoption
  - Phone Type
  - Scratch card top-ups

Reactions to Individual Shocks: Natural, Disease, Economic

• Sampled 1000 Rwandan Mobile Phone Users
  – Hashed IDs mapped to CDR
  – 30 questions to assess socioeconomic status and different shocks (sickness, drought, market crashes, deaths in the family, ...)

• Identifying Individual Behavioral Signatures
  – Deviations in
    • Mobility
    • Communication
    • Financial Transactions

Regional Ties & Cultural Norms: HIV/AIDS Knowledge, Attitudes, and Behavior

• Coupling CDR from 500k phones with DHS survey data in the Dominican Republic
  • Knowledge of HIV prevention
  • Misconceptions
  • Stigma
  • Higher-risk sexual behavior
Quantifying Slum Dynamics
Prediction of Ad-Hoc Housing Establishment

• How much of the growth of a city can be parameterized?
• Can we capture the dynamics of slums in a predictive model?

Universal Laws of Human Mobility?

- Are there common patterns in human movement across cultures and continents?
- How does season migrations impact computational epidemiological models of disease (malaria) eradication strategies?

SMS BloodBank

- Real-time monitoring of blood supplies at local district hospitals in Eastern Kenya
Surveys for Income Generation
txteagle: Mobile Crowdsourcing

- Mobile Crowdsourcing (mTurk)
- Tasks: translation, transcription, image tagging, surveys
- 15M East African can now earn income on their phones.

Empowering the largest knowledge workforce on Earth.

Reactions to Regional Shocks: Natural, Disease, Economic

- Spatial Dynamic Bayesian Anomaly Detection
  - Quantify tower-level behavioral priors

N = BILLIONS...

Bolivia, Dominican Republic, United States, China, Japan, Belgium, Thailand, Rwanda, United Kingdom, Kenya, Uganda, Saudi Arabia, Kuwait, India, Chad, Bahrain, Jordan, Lebanon, Brazil, Spain, Saudi Arabia, Indonesia, DRC, Gabon, Ghana, Ireland, Madagascar, Malawi, Niger, Nigeria, Sierra Leone, Argentina, Peru, Chile, Sudan, South Africa, Tanzania, Uganda, Zambia
thanks.

Aaron Clauset (The Santa Fe Institute)  Marta Gonzalez (MIT)
Stephen Guerin (RedFish)            Dirk Brockmann (Northwestern)
David Lazer (Harvard)               Marcel Fafchamps (Oxford)
Mika Raento (Google)                Edo Airoldi (Harvard)
Hannu Verkasalo (HUT)               Neil Ferguson (Imperial)
Peter Wagacha (University of Nairobi) Joshua Blumenstock (Berkeley)
John Quinn (Makerere)               Rob Claxton (British Telecom)
Cosma Shalizi (CMU)                  Michael Macy (Cornell)
Alessandro Rinaldo (CMU)             Alex (Sandy) Pentland (MIT)
Raja Hafiz (CMU)                     Ben Olding (Harvard)
Leon Danon (Warwick / Harvard)      Eric Horvitz (MSR)
Caroline Buckee (SFI / Oxford)       Luis Bettencourt (Los Alamos National Labs)
Sune Lehmann (Northeastern)          Jameson Toole (SFI)
Sir Lord Robert May (Oxford)         Amy Wesolowski (SFI)
Yves-Alexandre de Montjoye (Louvain) Harri Pennanen (Nokia)
Cris Moore (SFI / UNM)

Nathan Eagle
The Santa Fe Institute
nathan@santafe.edu