

# Die Fernerkundung des Internets: Neue Wege sozialwissenschaftlicher Datenerhebung

Nils B. Weidmann  
Universität Konstanz

*Tagung "Entscheidende Daten"*  
*Meersburg*  
*8./9. Mai 2014*

# Die digitale “Revolution” und die Sozialwissenschaften

- Mehr und detailliertere Daten: Vom Krieg zum einzelnen Konfliktereignis
- Dezentralisierte Informationsverbreitung: vom staatlichen Rundfunk zu Facebook
- Automatisierte Messung: von Umfragen zu Tweets

# Was bedeutet das für uns?

- Neue Messtechniken
- Neue Forschungsfragen

# Einfluss des Internet auf politische Mobilisierung

- “Befreiungstechnologie”:  
Kommunikationstechnologie fördert Protest und Demokratisierung.
- “Repressionstechnologie”:  
Kommunikationstechnologie fördert Überwachung durch staatliche Organe.

Quelle: time.com



Wael Ghonim

Quelle: guardianlv.com



Edward Snowden

# Empirische Untersuchung

- Experiment? Schwierig.
- Beobachtungsstudien: Wir müssen Internetnutzung über grosse Anzahl Fälle messen.
- Auf der substaatlichen Ebene, um Variation innerhalb von Ländern zu erfassen.

# Ein Messproblem

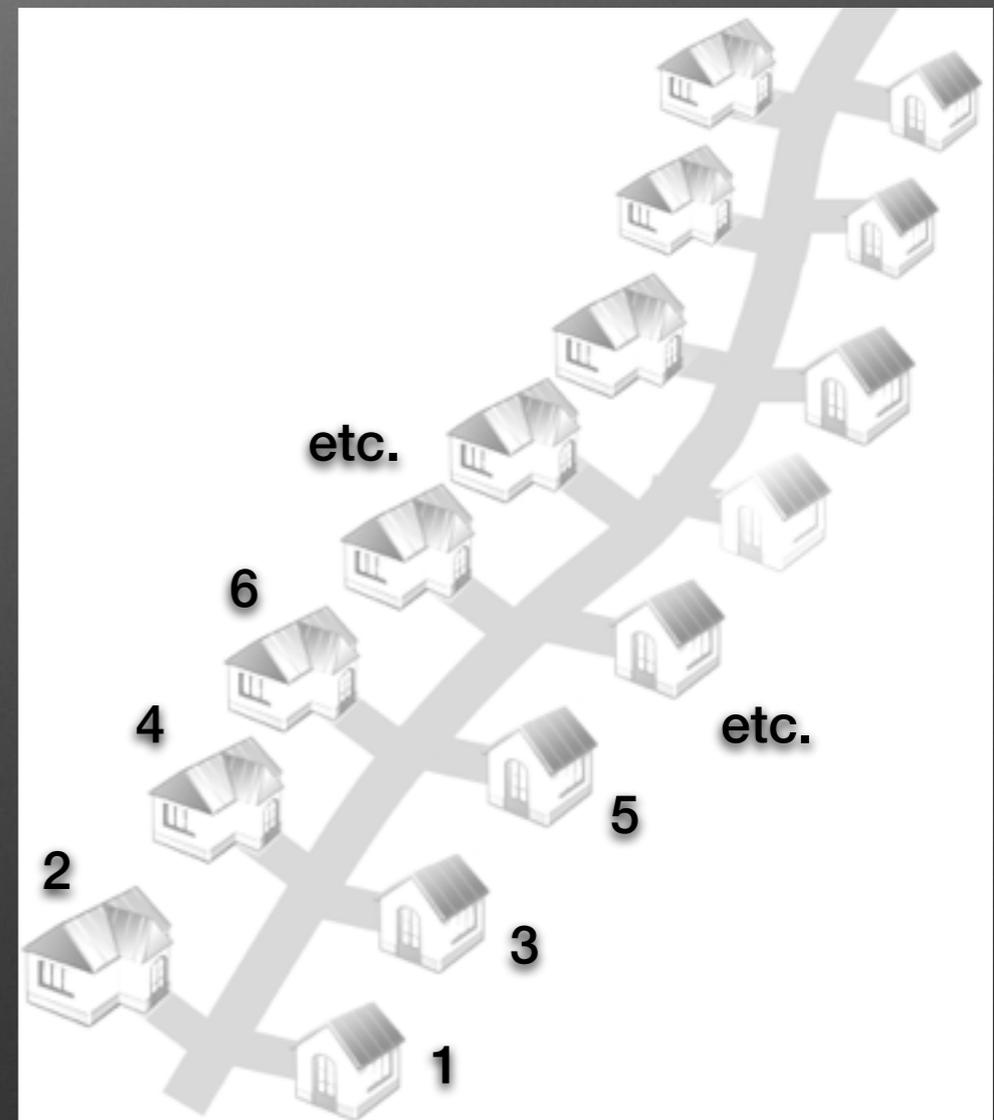
- Wie messen wir Internetnutzung...
  - ...auf substaatlicher Ebene
  - ...und für eine grosse (globale) Anzahl von Fällen?
- Staatliche Statistiken / Providerdaten / Umfragen: nur selten möglich
- Netzwerkmessung: Das Internet ist ein globales Netz, man kann Aktivität (teilweise) aus der Ferne beobachten

# Unser Ansatz

- Finde heraus, wo sich die Computer des Internet befinden...
- ... und zähle, wieviele dieser Computer sich in einer substaatlichen Einheit (z.B. Bundesstaat) befinden
- Probleme:
  1. Was sind die “Computer des Internet”?
  2. Wo befinden sie sich?

# Hintergrund: Das Internet in 2 Minuten

- Computer, die Datenpakete hin- und her senden
- Identifikation über Adressen: jedes Paket hat eine Quell- und eine Zieladresse
- Analogie: Hausnummern



# Problem 2: Wo befinden sich die Computer des Internet?

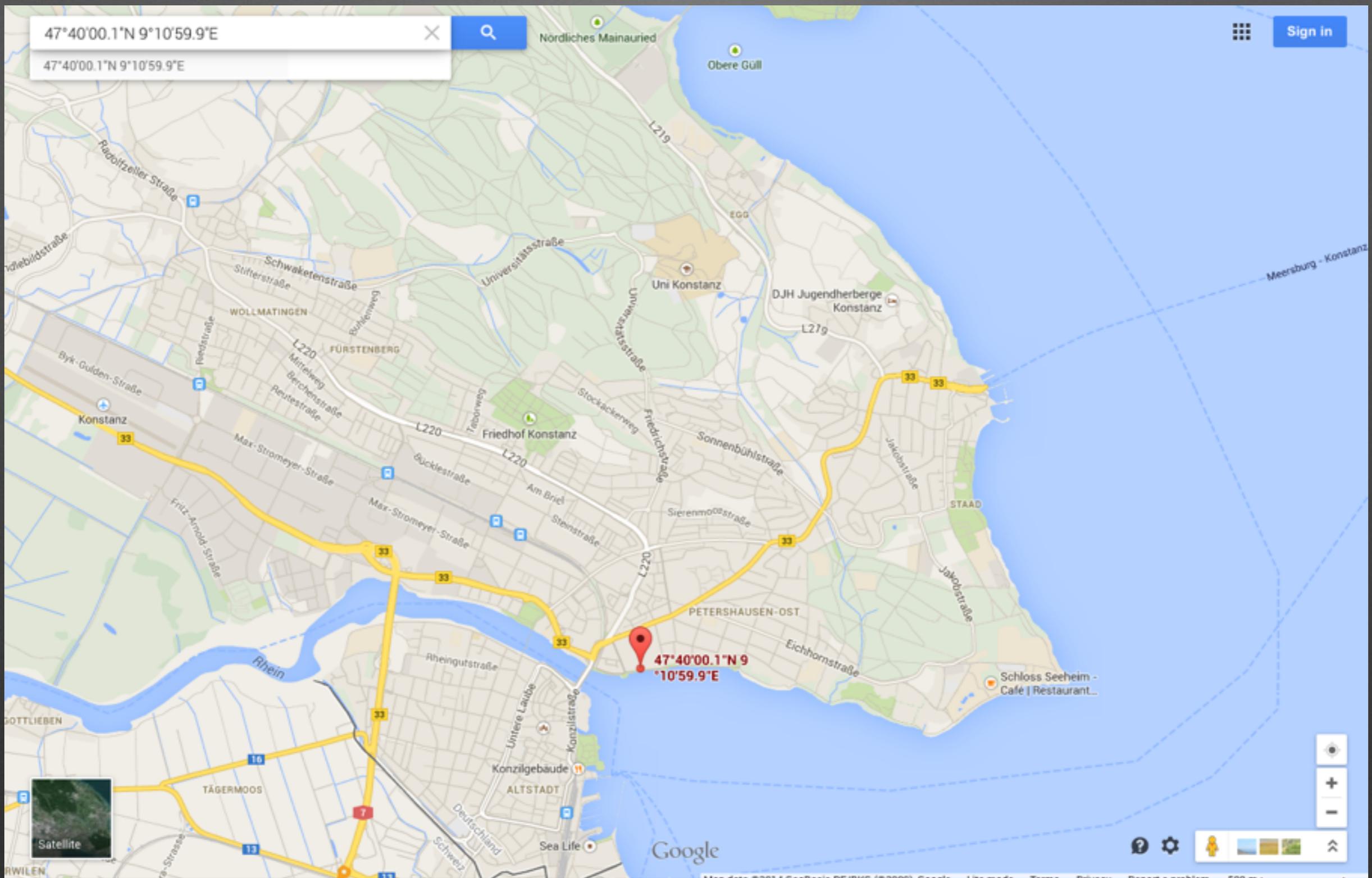
- Übersetzung Internetadresse in geographische Koordinaten durch Datenbanken
- Beispiel
  - Netzwerk der Universität Konstanz: 134.34.54.\*
  - Breite: 47.666698 Grad, Länge: 9.183300 Grad

47°40'00.1"N 9°10'59.9"E

47°40'00.1"N 9°10'59.9"E



Sign in



Quelle: maps.google.com

# Problem 1: Was sind die Computer des Internet?

- Wenn alle Internetadressen aktiv wären, könnten wir sie einfach aufzählen und lokalisieren
- Aber: Viele Adressen sind nicht aktiv
- In unserer Analogie: Viele unbewohnte Häuser / Hausnummern



Quelle: suedkurier.de

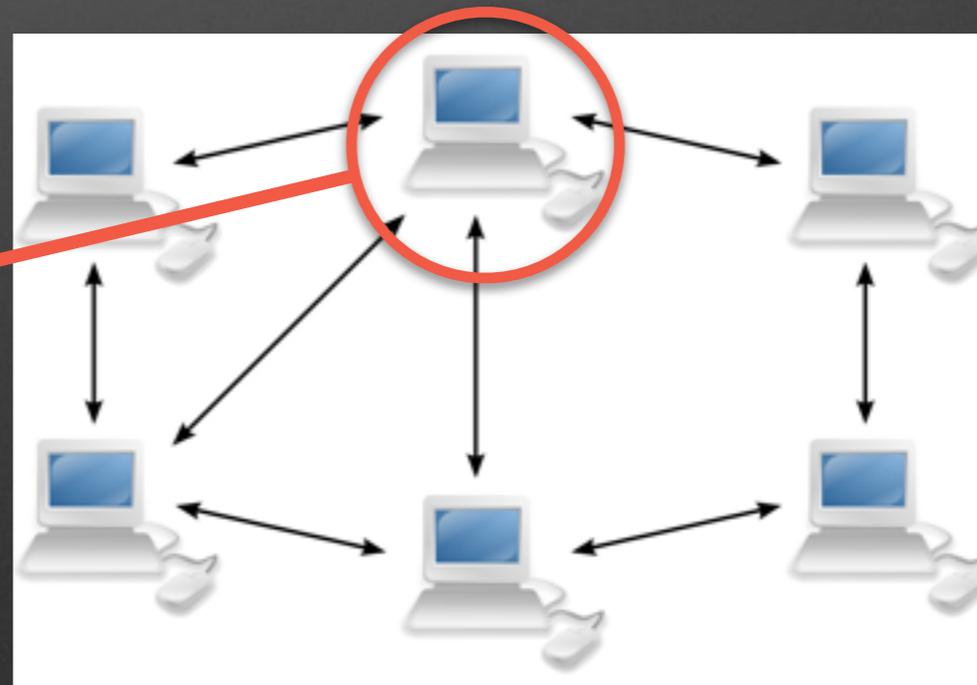


Quelle: buyhomesindetroit.com

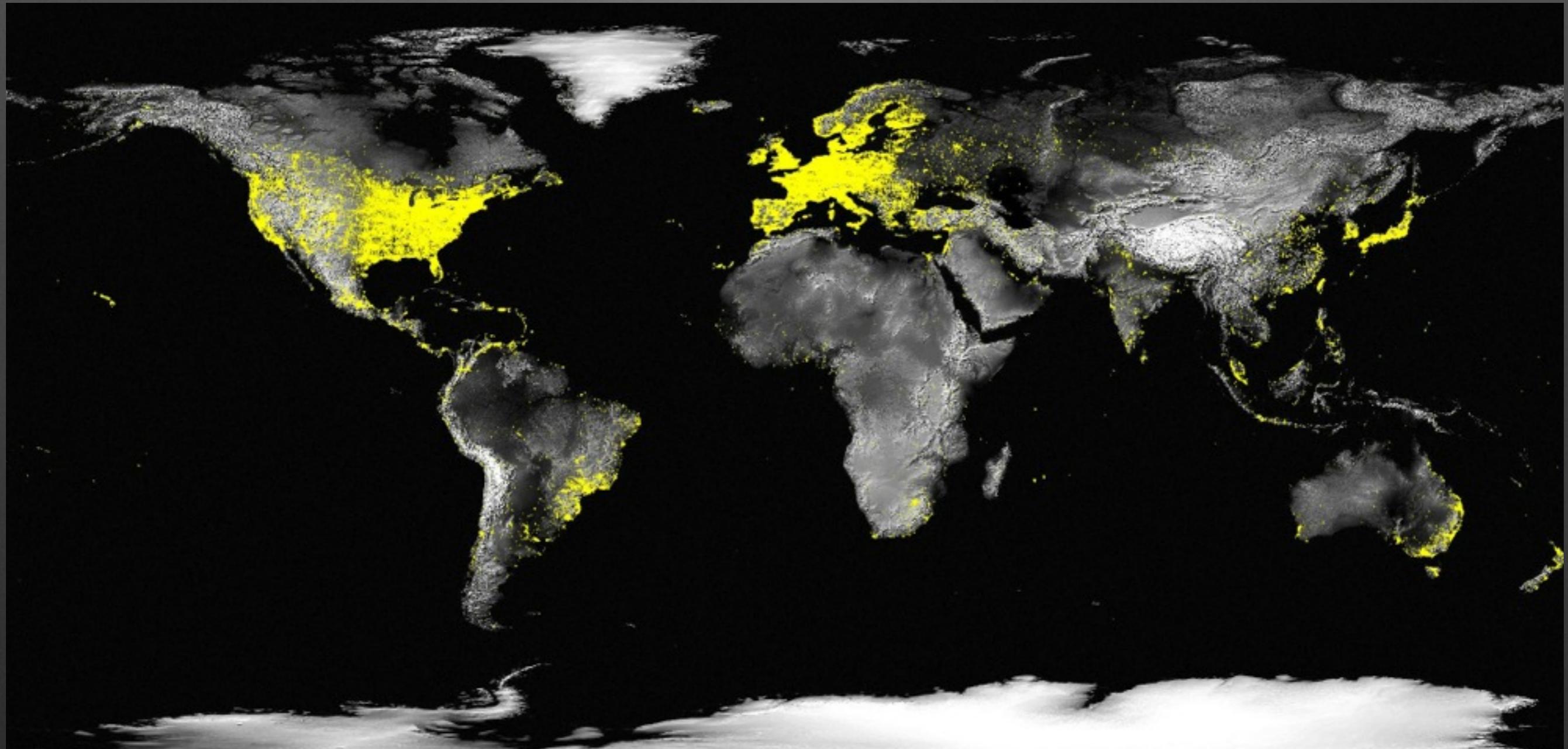
# Problem 1: Was sind die Computer des Internet?

- Die aktiven Adressen können durch “Fernerkundung” bestimmt werden
- Ansatz: Untersuchung des Datenverkehrs eines bestimmten Computers
- Passive Messung: Warten auf eingehende Verbindungen

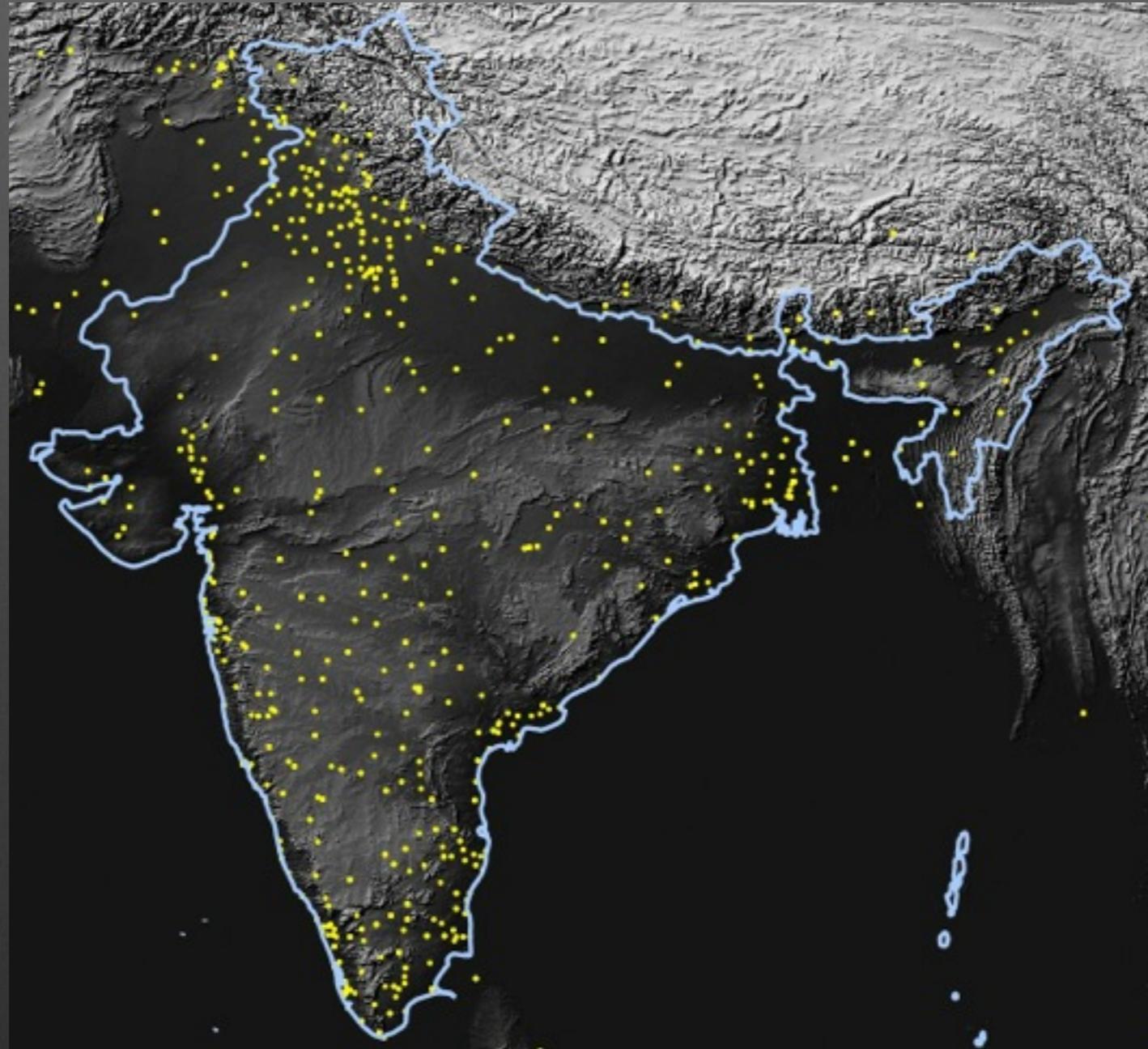
```
86.162.232.224 - - [17/Dec/2010:07:00:32 +0000] "GET /methodo  
86.162.232.224 - - [17/Dec/2010:07:00:32 +0000] "GET /resourc  
86.162.232.224 - - [17/Dec/2010:07:00:32 +0000] "GET /methodo  
86.162.232.224 - - [17/Dec/2010:07:00:32 +0000] "GET /resourc  
86.162.232.224 - - [17/Dec/2010:07:00:32 +0000] "GET /resourc
```



# Eine Weltkarte des Internet

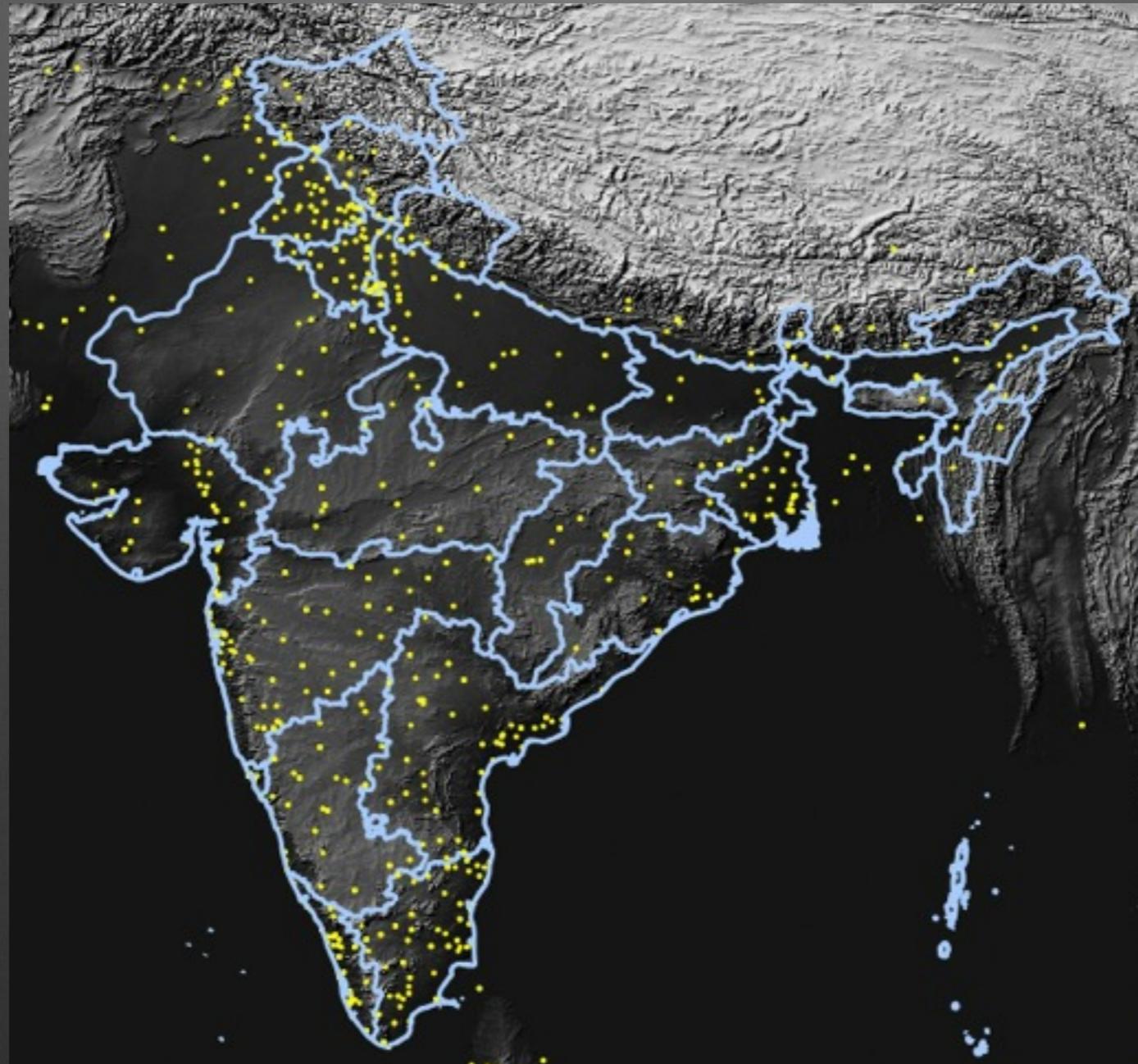


# Beispiel: Anzahl aktiver Adressen pro Land



Indien (2010)

# Beispiel: Anzahl aktiver Adressen pro Bundesstaat



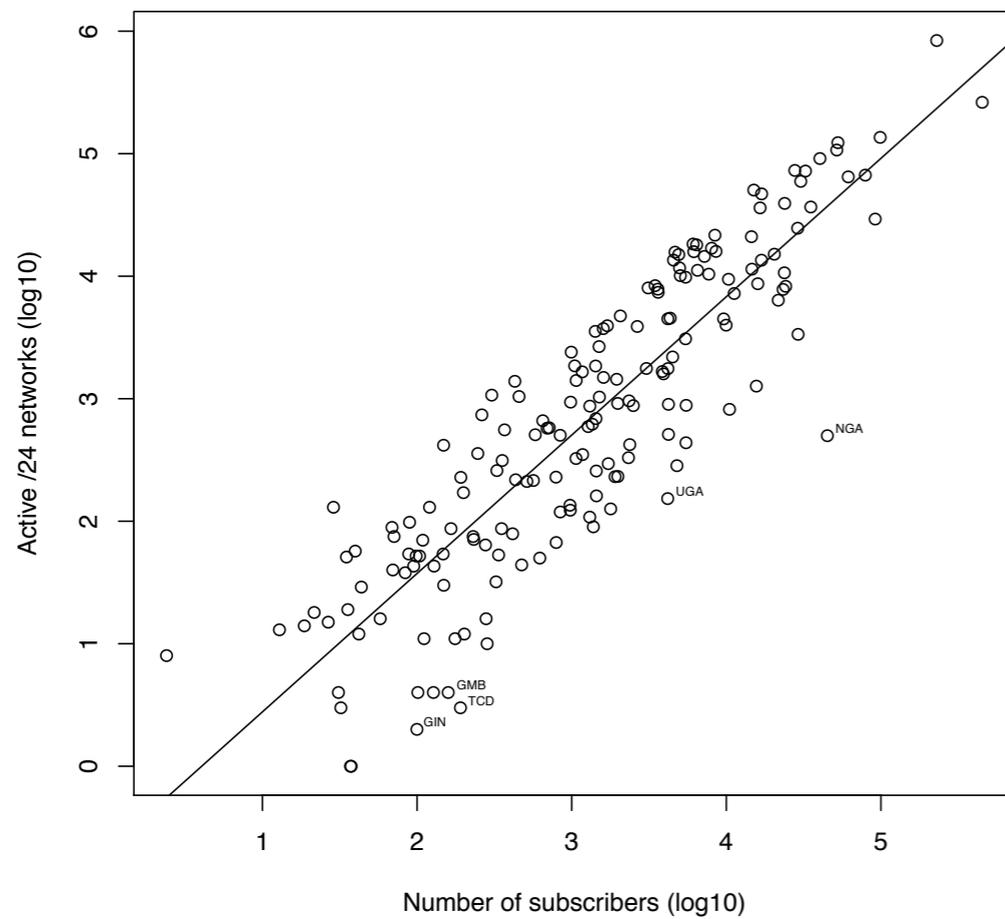
Indien (2010)

# Kann das überhaupt funktionieren?

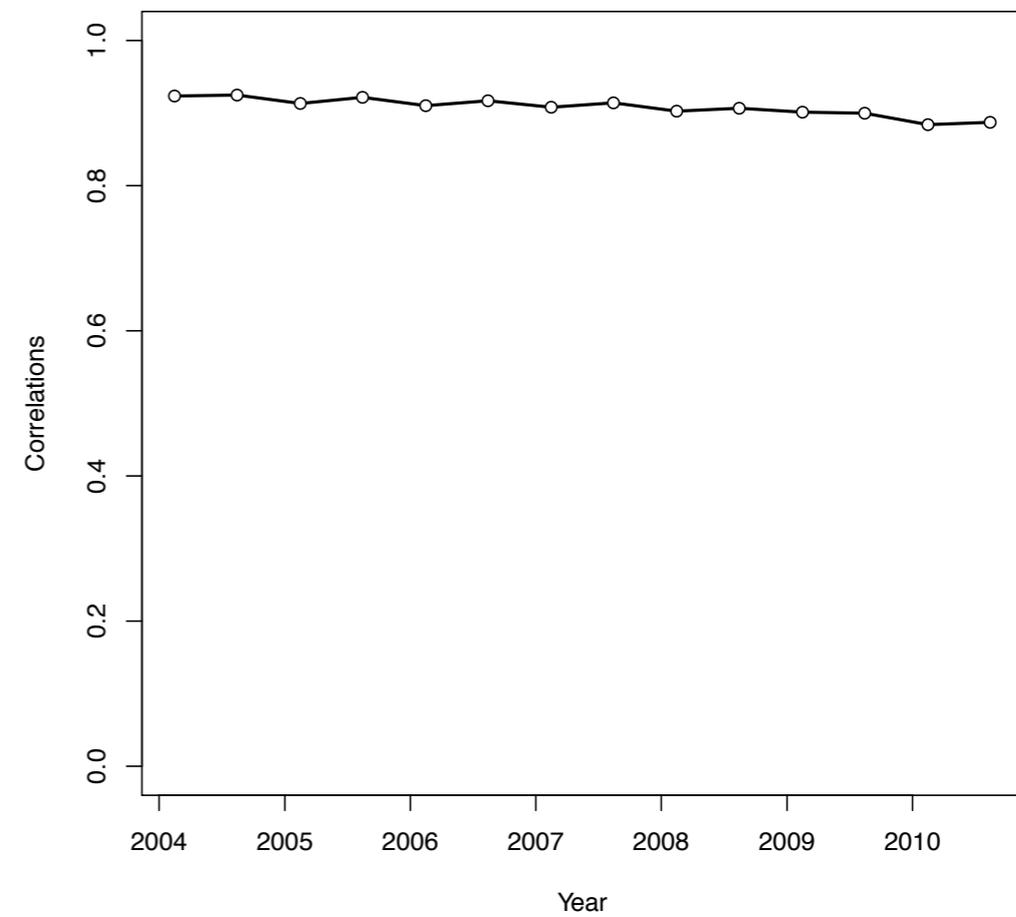
- Wir verwenden die Anzahl aktiver Computeradressen zur Schätzung der Menge der Internetnutzer.
- Wir verwenden eine “Schweizer Sicht” aufs Internet, indem wir nur Datenverbindungen berücksichtigen, die durch einen grossen Schweizer Provider gehen.

# Ergebnisse: Ländervergleich

ITU Estimates vs. Active Networks (August 2010)  
 $r=0.887$

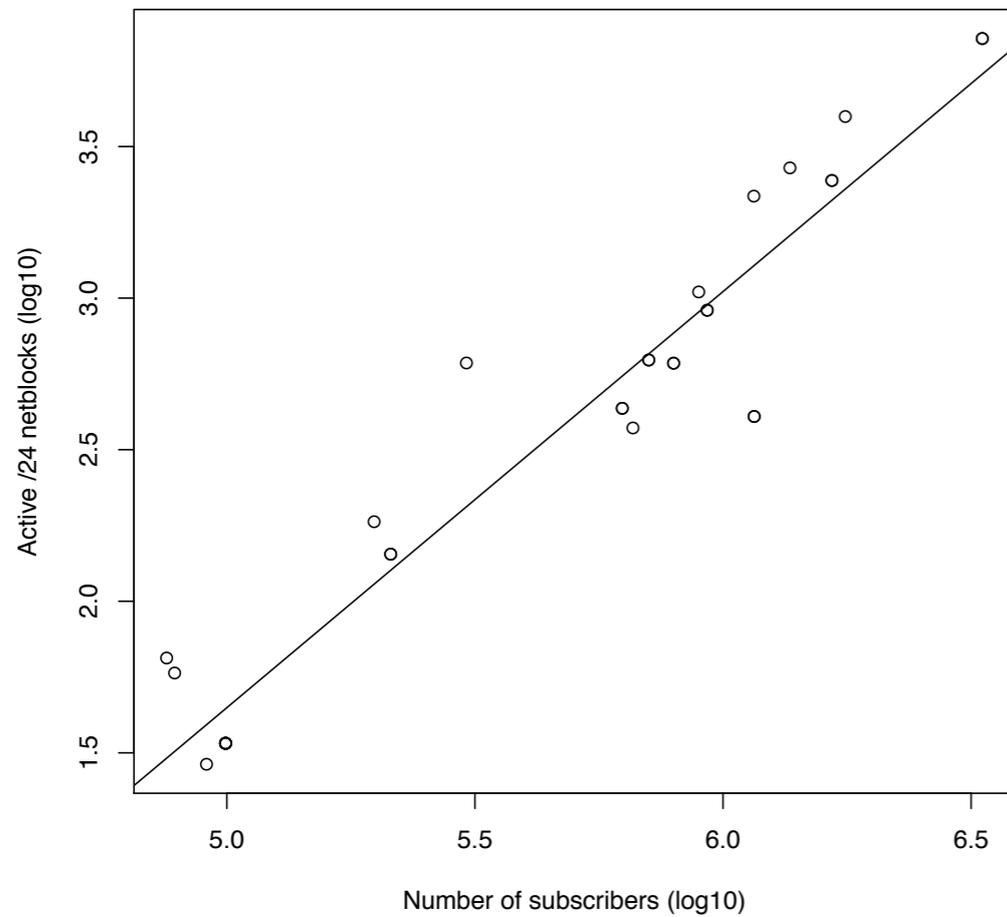


ITU Estimates vs. Active Networks (2004 - 2010)

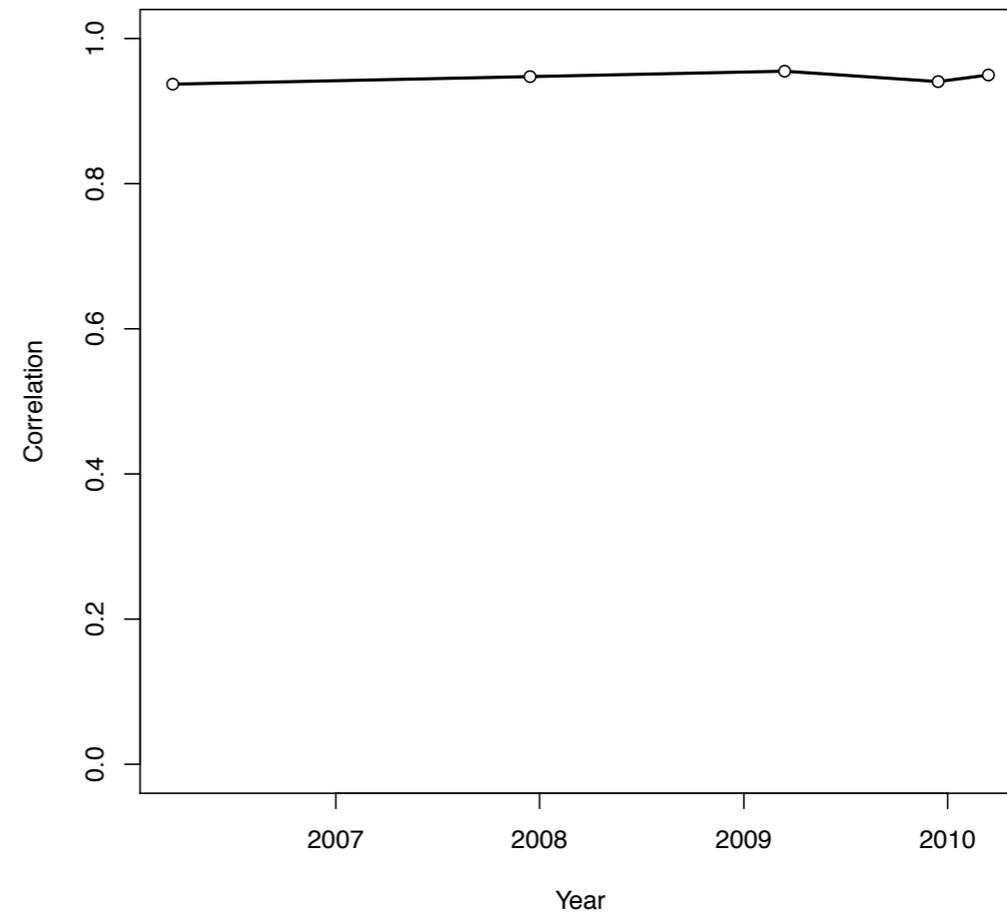


# Ergebnisse: Indien

India: Official Estimates vs. Active Networks (March 2010)  
 $r=0.961$



India: Official Estimates vs. Active Networks (2006–2010)



# Fortschritt?

- Globaler Datensatz der Internutzung über die Zeit, mit (begrenzter) substaatlicher Auflösung
- Vielfältige Anwendungsmöglichkeiten: Politische Mobilisierung, politische Gewalt, Entwicklung, staatlicher Ausbau von Infrastruktur
- “Big Data”! Insgesamt 38.019.137 Einträge in unserer Datenbank (2004-2013)
- Aber:
  - Keine Information über Inhalte (was wird übermittelt?)
  - Zu grobe Auflösung, um Aussagen über die Individualebene zu machen

# Zum Schluss

- *Gracias*: Suso Baleato für hervorragende Arbeit an diesem Projekt
- Ευχαριστώ / *Merci vielmol* / *Thank you*: Fontas Dimitropoulos, Edy Glatz und Brian Trammell (ETH Zürich)
- *Dankeschön* an die Alexander von Humboldt-Stiftung für finanzielle Unterstützung (Sofja Kovalevskaja Preis)

Unterstützt von / Supported by



Alexander von Humboldt  
Stiftung / Foundation