Complexity Science and Quantification of Social Science

Jianbo Gao

Institute of Complexity Science and Big Data Technology

Guangxi University

jbgao.pmb@gmail.com, jianbo_gao_gxu@163.com

http://www.gao.ece.u.edu/, http://bigdata.gxu.edu.cn/

March, 2016, IPAM, UCLA

Prelude

- "I have a lot of data. What can I do?"

  --- degree of distress proportional to amounts of data?
- In social science, the most interesting is the change
- Statistical analysis, data mining, machine learning are all useful, but not sufficient

  --- In business, making money is the sole purpose; in science, a small error could cost enormously
- Complexity science offers another solution
  - Focuses more on universally applicable concepts than models: nonlinearity, bifurcations, sensitive dependence on small disturbances, fractal self-similarity, long memory, extreme variations, nonstationarity, etc
  - Most interested in changes and causal relations
- Human brain thinks logically or causally

My background

- Education & experience:

  Automatic control, fluid mechanics, meteorology, electrical engineering at UCLA (communications and networking), bioinformatics, healthcare, econometrics, political science

- Real focus:

- Self-taught nonlinear dynamics, obsessed by the different foundations of deterministic chaos theory and random fractal theory (eventually find a solution to synthesize them)
- Increasingly attracted to develop multiscale methods to solve important real world problems than to engage purely theoretical research
- Believe data analysis is the first step --- a truly significant insight will always lead to more questions

Ongoing projects in our Institute

- Biomedical signal analysis (hearing testing, apnea, Alzheimer's disease, etc)
- Air pollution
  - spatial-temporal dynamics of PM2.5 pollution
  - persistency and long-range correlation
  - season-dependent diurnal variations
- Identification of cheating behavior in online bank transactions
  - Different operations are governed by different math laws
- Forewarning of financial crises and policy shocks
- Quantification of social science

Introduction to fractal and multifractal

- A part is (exactly or statistically) similar to another part, or the whole — scale-free
- Clouds; mountains; trees; etc. (Images: not computer-made, but photos of Jiu Zhai Gou)
- Power-law relation — a straight line in a log-log plot (scaling law)
- Power-law relation is the origin of self-similarity
- Many (or possibly infinitely many) power-law relations — Multifractal

Scope of fractal theory

- Fractal geometry — both deterministic and stochastic
- Chaotic attractors often have fractal structure — deterministic
- Dynamical random fractal theory
  - $1/f^{\alpha}$ processes
  - A subclass, $1/f^{2H+1}$ process, where H is called Hurst parameter, has long-range correlations or long memory
    - ⋆ Depending on whether $0 < H < 1/2$, $H = 1/2$, or $1/2 < H < 1$, the process is said to have anti-persistent, short memory (or memoryless), and persistent long-range correlations
  - Multiplicative cascade multifractals
  - Levy processes
- Chaos and random fractals have different foundations — incorrect to use chaos to "mean" both chaos and fractal!
- Applications: cyber security, financial crises, river flow dynamics, political instability

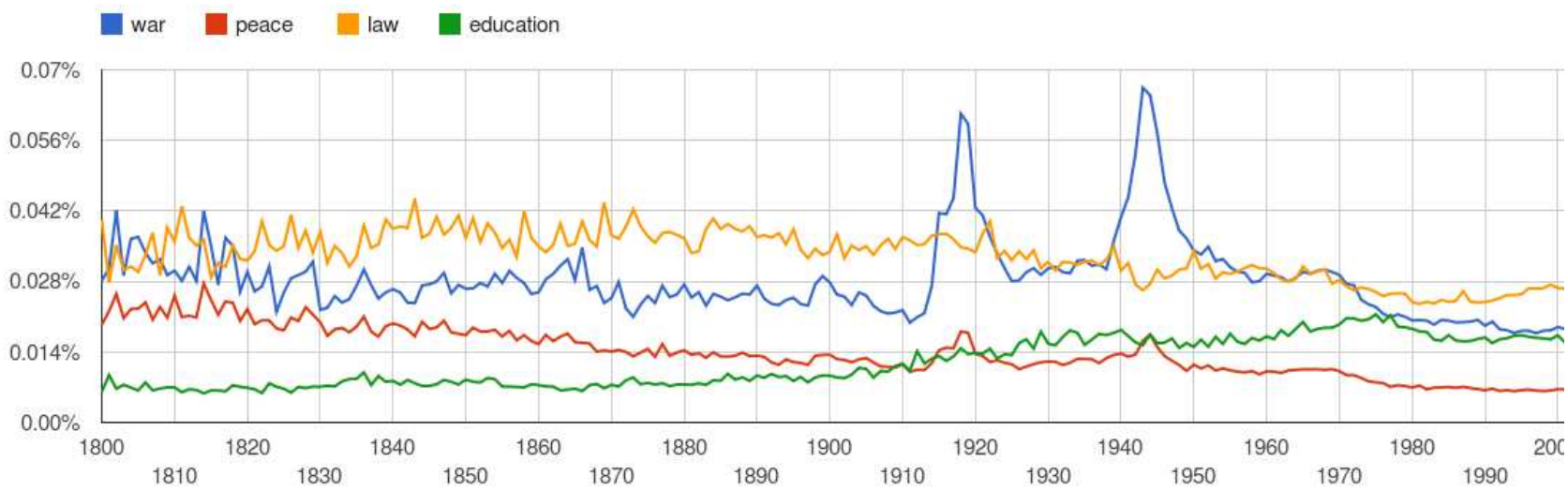
Google books Ngram Viewer

Graph these case-sensitive comma-separated phrases:
war,peace,law,education

Tweet 0

between 1800 and 2008 from the corpus English with smoothing of 0.

Search lots of books



Search in Google Books:

1800 - 1819	1820 - 1874	1875 - 1892	1893 - 1979	1980 - 2008	law	English
1800 - 1812	1813 - 1825	1826 - 1875	1876 - 1969	1970 - 2008	peace	English
1800 - 1817	1818 - 1942	1943 - 1956	1957 - 1978	1979 - 2008	war	English
1800 - 1835	1836 - 1965	1966 - 1978	1979 - 1993	1994 - 2008	education	English

Run your own experiment! Raw data is available for download here.

Human response to natural phenomena: $1/2 < H < 1$

1-grams	Hurst Parameter (H)	
	AFA	DFA
avalanche	0.63 ± 0.06	0.79 ± 0.06
comet	0.60 ± 0.03	0.73 ± 0.04
drought	0.81 ± 0.05	0.69 ± 0.09
earthquake	0.65 ± 0.02	0.72 ± 0.03
erosion	0.85 ± 0.06	0.86 ± 0.08
fire	0.67 ± 0.05	0.70 ± 0.03
flooding	0.85 ± 0.06	0.72 ± 0.08
hurricane	0.70 ± 0.03	0.69 ± 0.08
landslide	0.66 ± 0.05	0.41 ± 0.20
life	0.62 ± 0.03	0.65 ± 0.06
lightning	0.63 ± 0.03	0.70 ± 0.03
mudslide	0.80 ± 0.02	0.58 ± 0.28
tornado	0.59 ± 0.02	0.64 ± 0.06
tsunami	0.81 ± 0.05	0.66 ± 0.03

Human response to social phenomena: $H > 1$ —nonstationary

1-grams	Hurst Parameter (H)	
	AFA	DFA
christian	0.85 ± 0.05	0.95 ± 0.08
communism	1.32 ± 0.04	1.44 ± 0.05
crisis	1.15 ± 0.05	1.13 ± 0.08
democracy	1.18 ± 0.02	1.07 ± 0.07
education	1.04 ± 0.05	1.09 ± 0.13
environment	1.13 ± 0.04	1.24 ± 0.08
famine	0.74 ± 0.02	0.66 ± 0.06
malnutrition	1.10 ± 0.07	1.08 ± 0.11
politics	1.14 ± 0.03	0.99 ± 0.06
population	1.01 ± 0.06	0.98 ± 0.10
recession	1.33 ± 0.05	1.06 ± 0.07
socializing	1.28 ± 0.07	1.28 ± 0.09
stock	1.01 ± 0.05	0.99 ± 0.11
unemployment	1.32 ± 0.04	1.28 ± 0.04

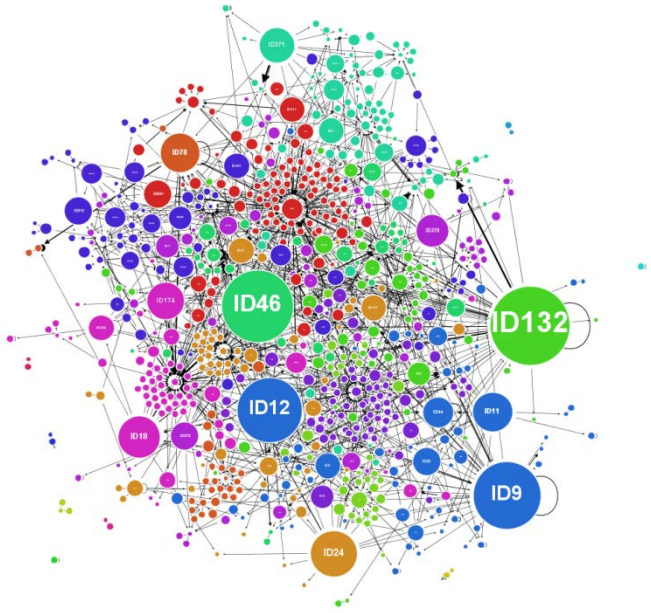
Insights from human response to natural and social phenomena

- Gao et al. 2012 Journal of Royal Society Interface: analyzed yearly usage frequencies of specific words that describe social and natural phenomena, as derived from 5 million books published over the course of the past two centuries
- Find that human response to social and natural phenomena are governed by fundamentally different processes
  - Response to natural phenomena: like noise processes involving individuals
  - Response to social phenomena: like random walk processes involving mass (integration of individuals)
- Usage of survived & extinct irregular verbs also like mass and individual dynamics (unpublished)
- Speculation: is formation of ideas, social moods, etc., governed by similar principles?

Influence propagation on cyberspace

- Nanjing plane tree protection: protest Nanjing Government for transplanting thousands of plane trees

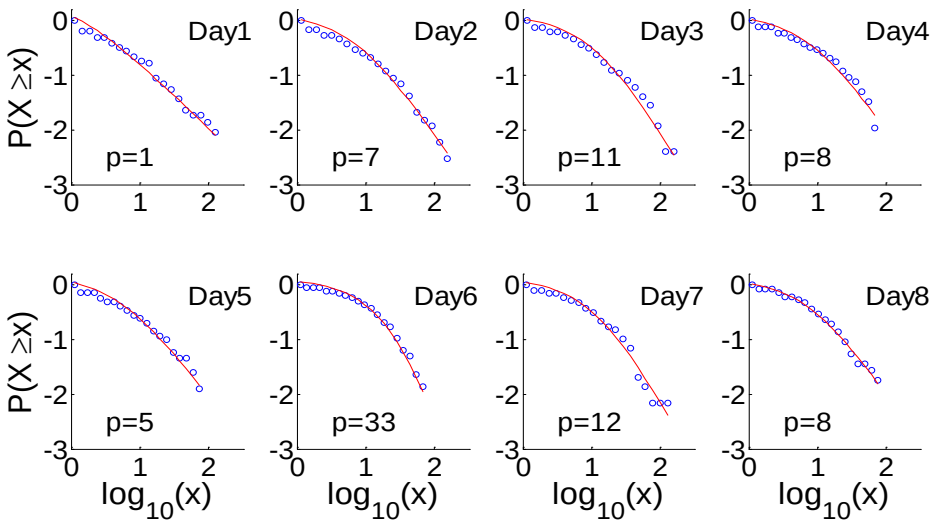
- Event's time and space: Sina Weibo microblog 2011.3.14-3.21



- Social networks: reposting/being reposted microblogs

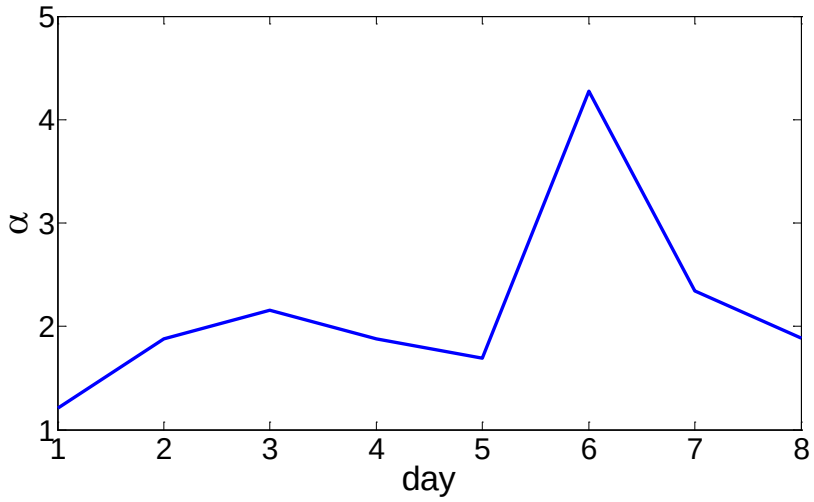
- Data: 16253 repostings (2483 nodes of users, 11976 edges of relations).

Influence of blogs follows Omori's law



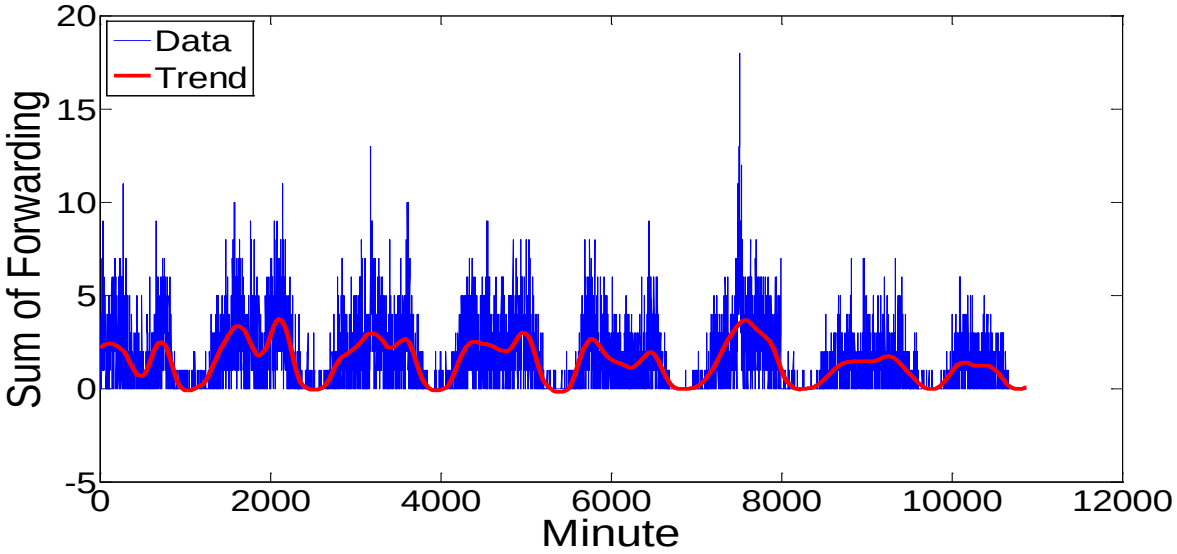
Degree distribution of daily reposted networks follows Omori's law describing the distribution of after-shocks following a major earthquake

$$P[X \geq x] = \left(\frac{1}{x + p} \right)^{\alpha}$$



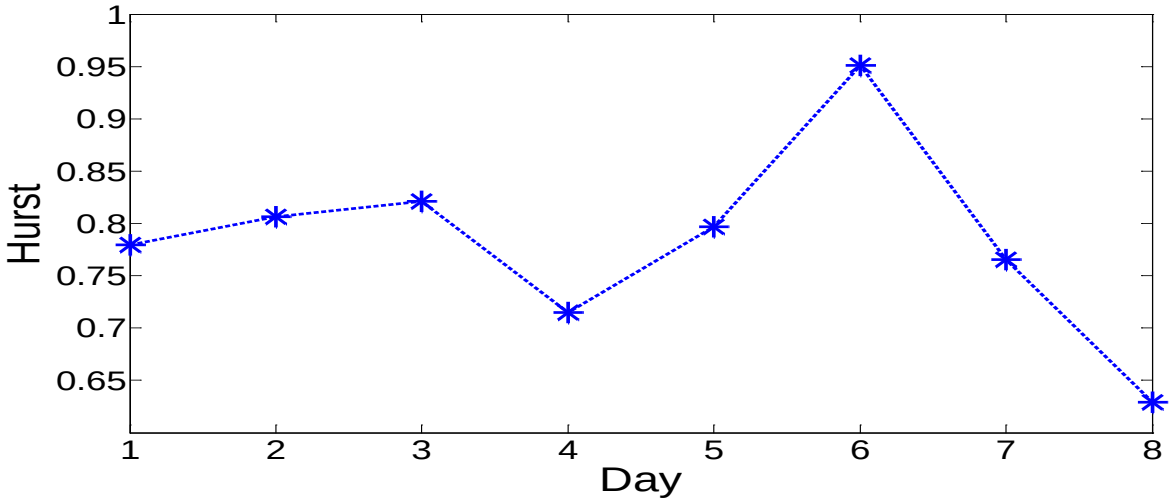
Variation of α with time, with its value reaching maximum on the 6th day when the mass parade of protest occurred

Consensus of opinion among activists is characterized by the Hurst parameter



◆Temporal variation of the network's activity

◆Variation of the Hurst parameter with time

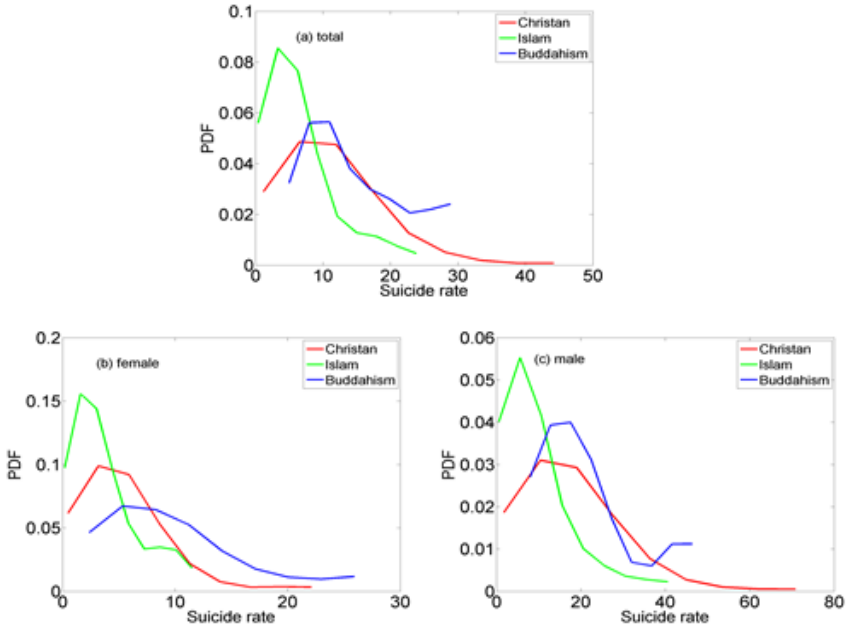


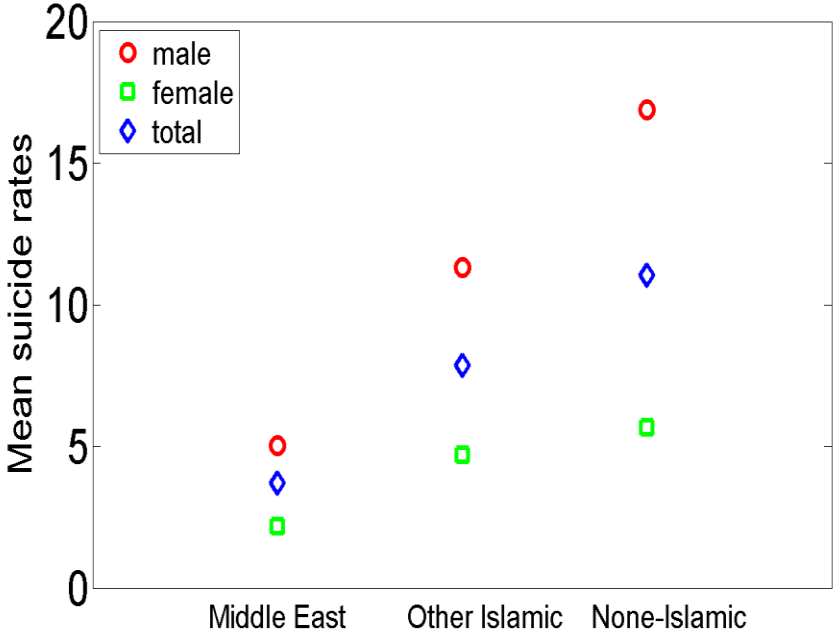
Religious culture, Suicide, and Terrorism: A Global Quantitative Analysis

D.F. Yu, J.B. Gao, H.L. Huang,

H.S. Feng, F.Y. Liu

Results





- Suicide rate: Buddhist > Christian > Islamic
  - Within each religious culture: Male > Female
- Islamic suicide rates: Middle East < Other Islamic < None-Islamic
  - coincidence: terrorism is the most violent in Middle East

Further problems

- Quantitatively, how different are the structures of the networks associated with different religions?
  - Exactly how integration and regulation are provided?
- What provides an integrative and regulative function in China?
  - (government, heresy, etc)

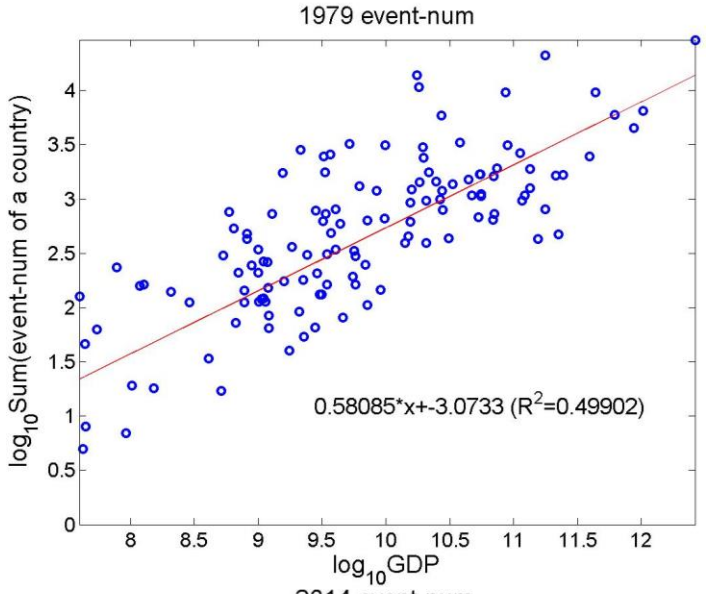
Quantitative social science:

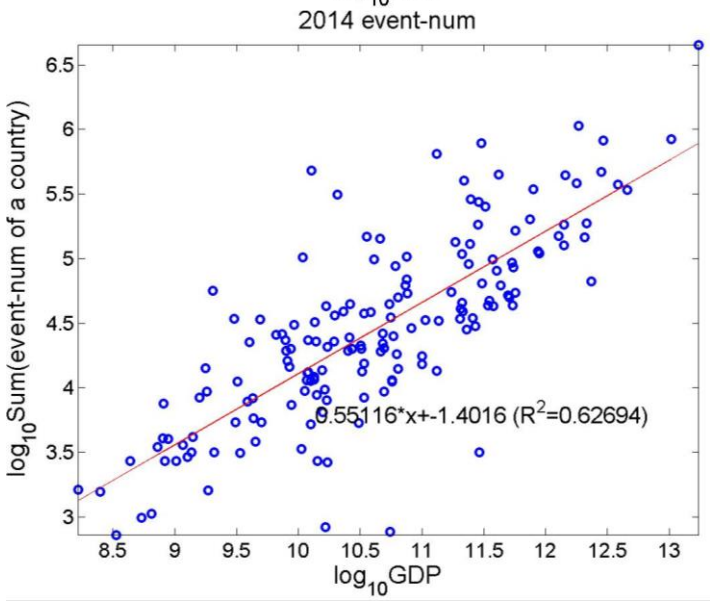
- Integrated Crisis Early Warning System (ICEWS)
  - Purpose: to forecast and respond to crises
  - Support: DARPA
    - Event data collected only if relevant to crises
- A much larger open data source: GDELT (global database of event, language, and tone)
  - Essentially all events reported in media are collected
  - Opportunities: do a lot more than ICEWS purports to do
  - Challenges: Huge number of events, some erroneous, are mixed together

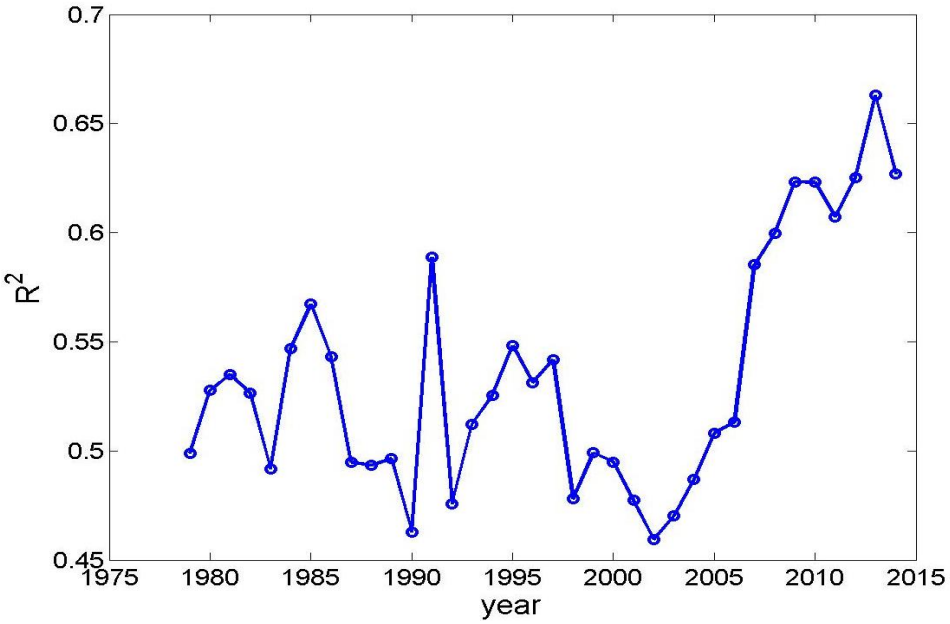
Political event data

- The event data set analyzed here is called Global Database of Events, Language, and Tone (GDELT)
- GDELT events are drawn from a wide array of news media, both in English and non-English, from across the world, ranging from international to local sources in nearly every country
- GDELT includes more than 300 million unique events across all countries, during the period from 1979 to the present
- These data were produced by the TABARI automated coding software (http://eventdata.psu.edu/software.dir/tabari.html) using the CAMEO event and actor coding system
- Basic structure of event data:
  - Actor 1 interacts with actor 2
  - Who and where Actor 1 & 2 are
  - When
  - Event score (Goldstein scale: $\pm$ for good & bad; $\sim$ 300 events)
  - ...

Potential of GDELT: Number of Events vs. GDP





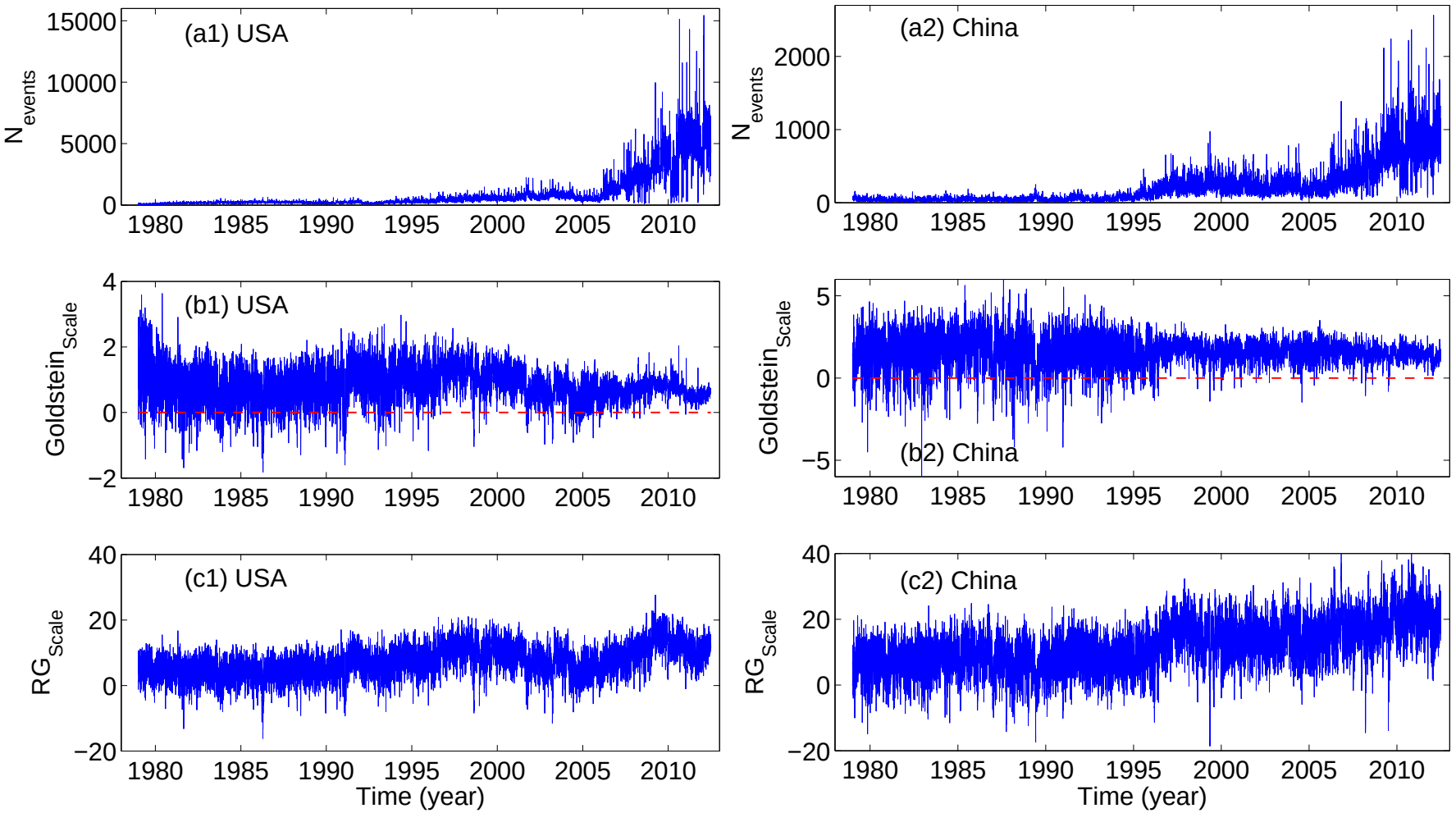


- A positive correlation also exists between the Event-num and GDP of almost all the countries in the world.

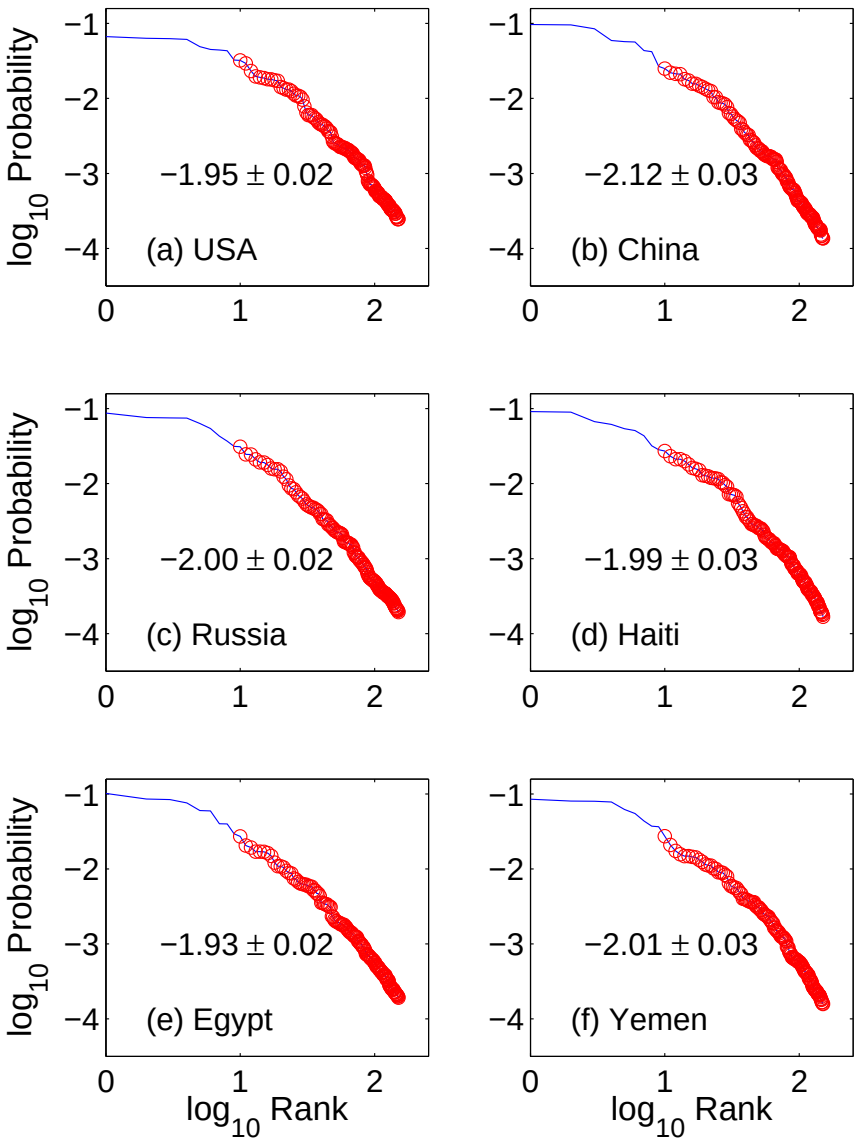
✓ All above mean that media reports of a country is related with its economy.

Daily event totals for USA and China

- Number of events grows exponentially
- Daily mean of the Goldstein-scale intensity data are nonstationary



Distributions of events ordered by occurrence frequency



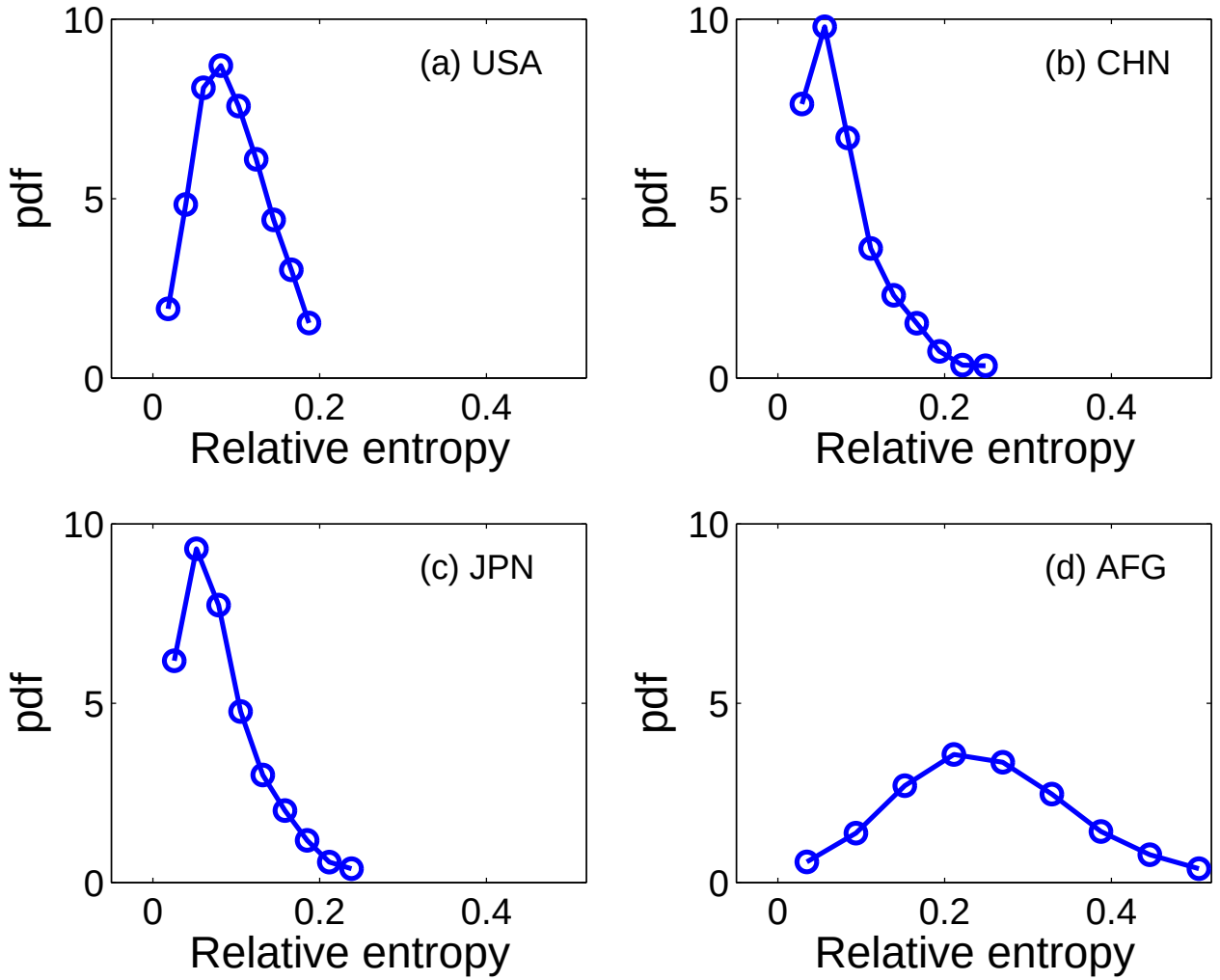
- Zipf-Mandelbrot law

$$p(k) = c/(k+q)^{\alpha}$$

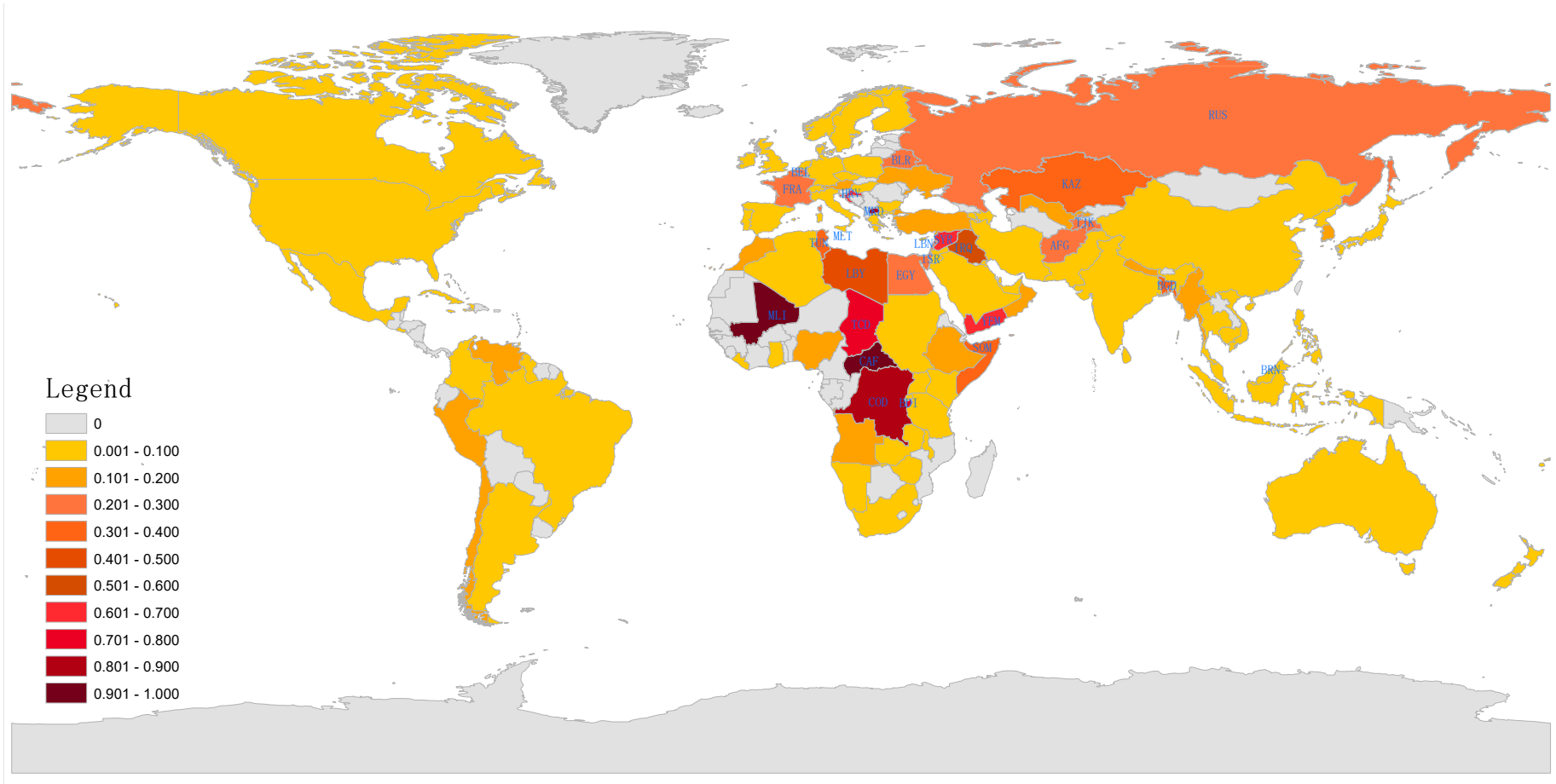
  characterizes both
  - events for the whole world in a short period of time
  - events in a specific country during the past 30+ years
- α only varies slightly among ~ 50 countries examined
  $\longrightarrow$ quasi-universal law
- Regardless of its size and political influence, a country engages in a set of activities well described by the CAMEO event code book

Quantifying the difference between the activities of 2 countries by relative entropy (or Kullback-Leibler distance)

$D(p||q) = \sum_i p_i \ln\left(\frac{p_i}{q_i}\right)$

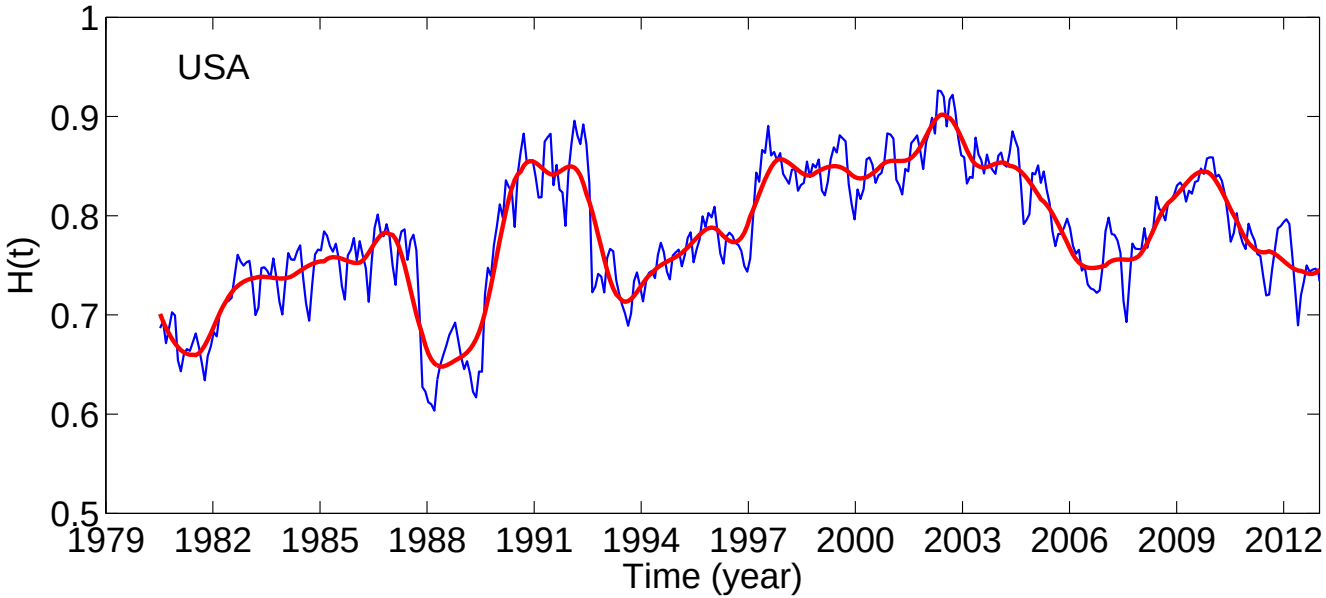


Global political "hotness" map for November,2015



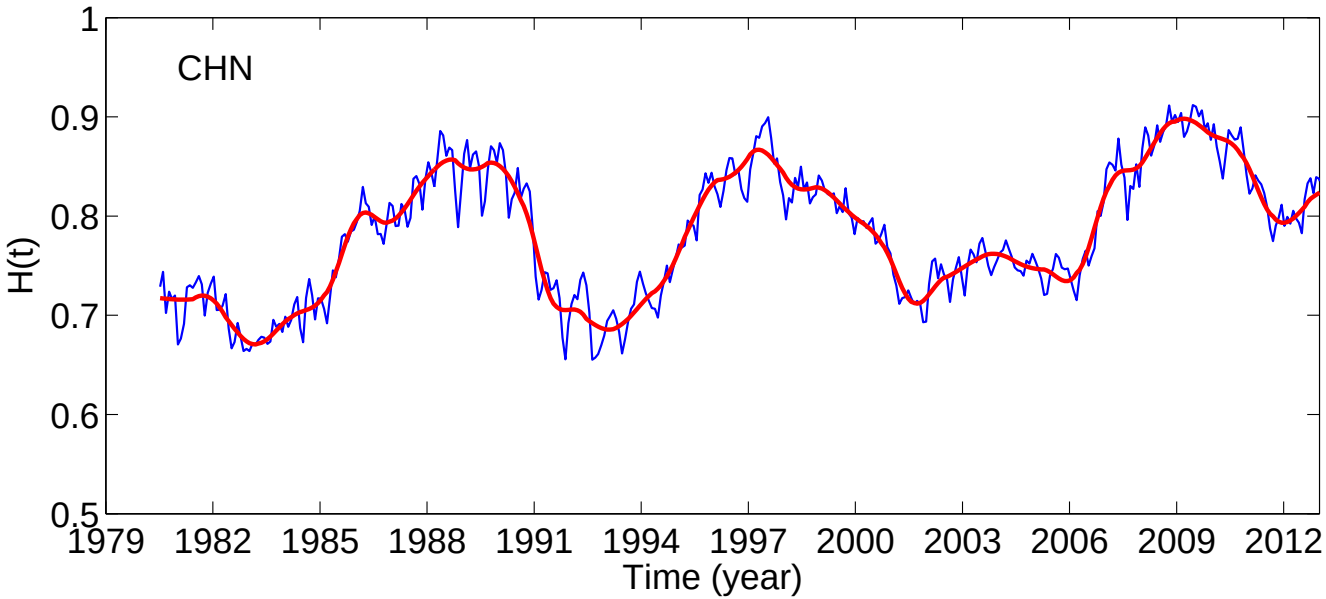
Evolution of political course in USA

- Temporal evolution: 1 month & 1 year for blue and red curve
- Three large decreases in $H(t)$:
  - Two can be easily associated with the two Iraq wars
  - The most interesting is the first sharp drop in $H(t)$ that occurred around 1987 , suggesting that the cold war between USA and the Soviet Union also had greatly strained USA



Evolution of political course in China

- The most eminent variation in $H(t)$, around 1990, was caused by the infamous Tianmen Square event
- A gradual decrease in $H(t)$, occurred around 1997, when Deng Xiaoping died
- Since then, $H(t)$ continued to decrease, coinciding with a lot of hectic events,including banning of China Democracy Party in 1998, banning of Falun Gong in 1999, NATO bombing of the Chinese embassy in Belgrade, Yugoslavia in 1999, and Tianmen Square self-immolation incident by Falun Gong members in 2001
- Local minimum of $H(t)$ around 2002, from Jiang Zhemin to Hu Jintao
- $H(t)$ increases during 2 economic reforms, starting around 1982 & 1992



Quantitative study of geopolitical relations through GDELT analysis

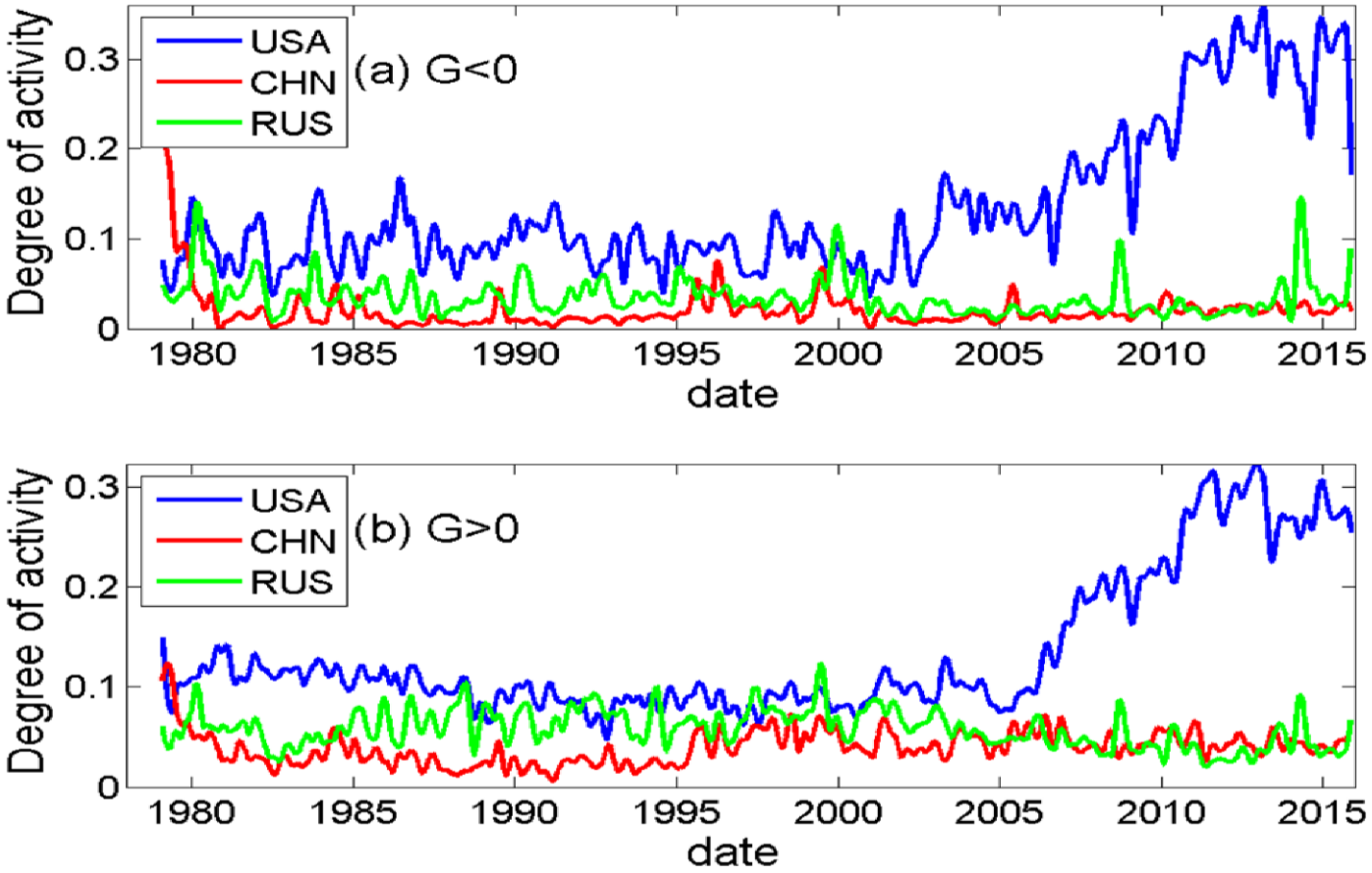
- An ideal direction for achieving excellency
- Let A be a matrix characterizing the daily relations among all the countries, where the element a_{ij} can be the total daily no. of events between countries i and j, or the total Goldstein scale (or the total positive or negative G-scale), etc.
- Singular value decomposition (SVD) of A:

$$A = U\Sigma U^T$$

  where $\sigma_1 \geq \sigma_2 \geq \cdots$
- Project the activity of country i onto the base vectors of U to obtain $(\alpha_1^i, \cdots, \alpha_n^i)$
- A measure of the activity ($\sim$ strength?) of country i:

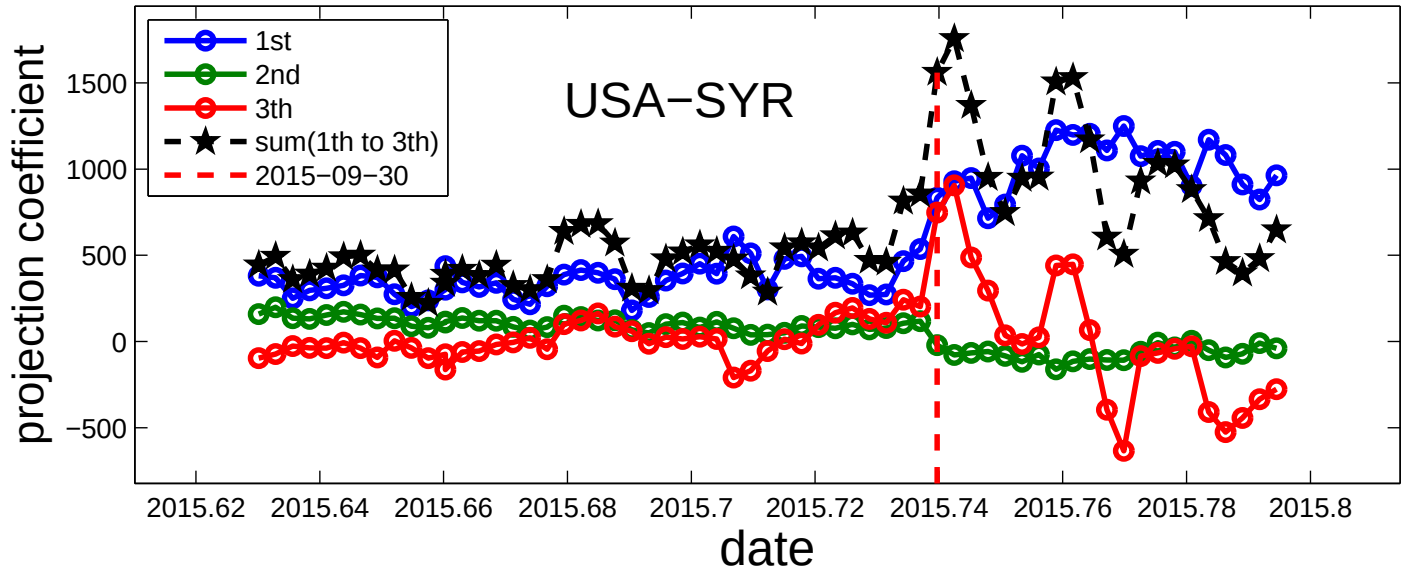
$$S_i = \sum_{k=1}^{n} |\alpha_k^i| \sigma_k^2$$

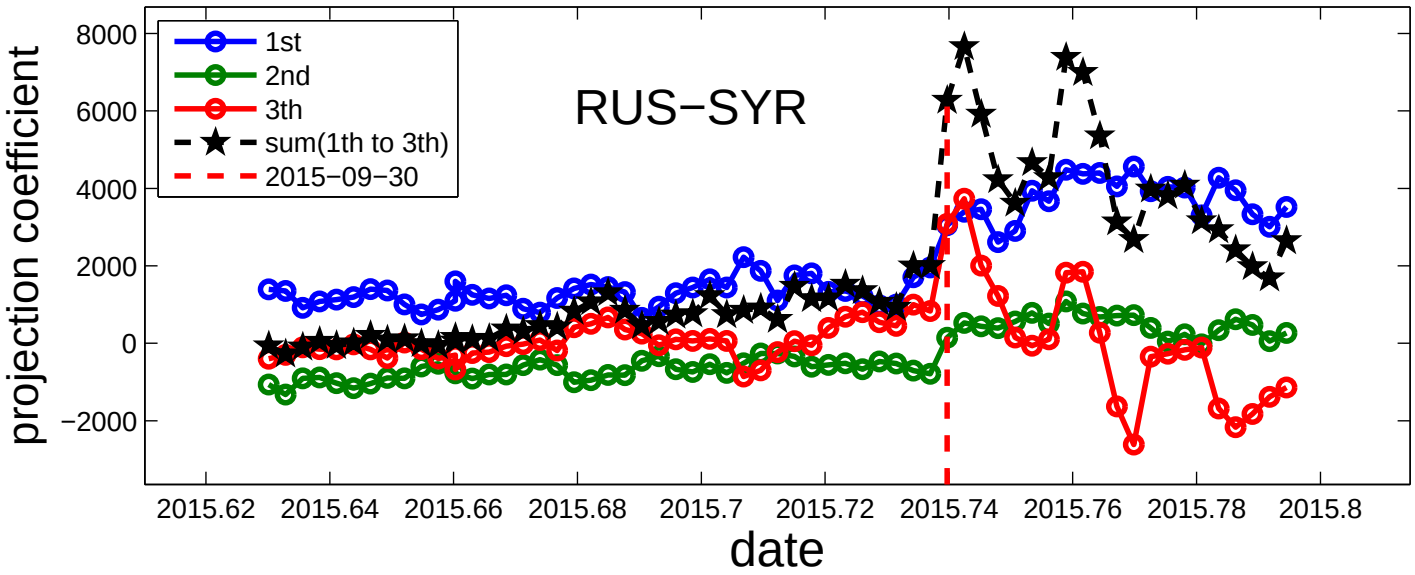
Activity Degree for 3 Major powers



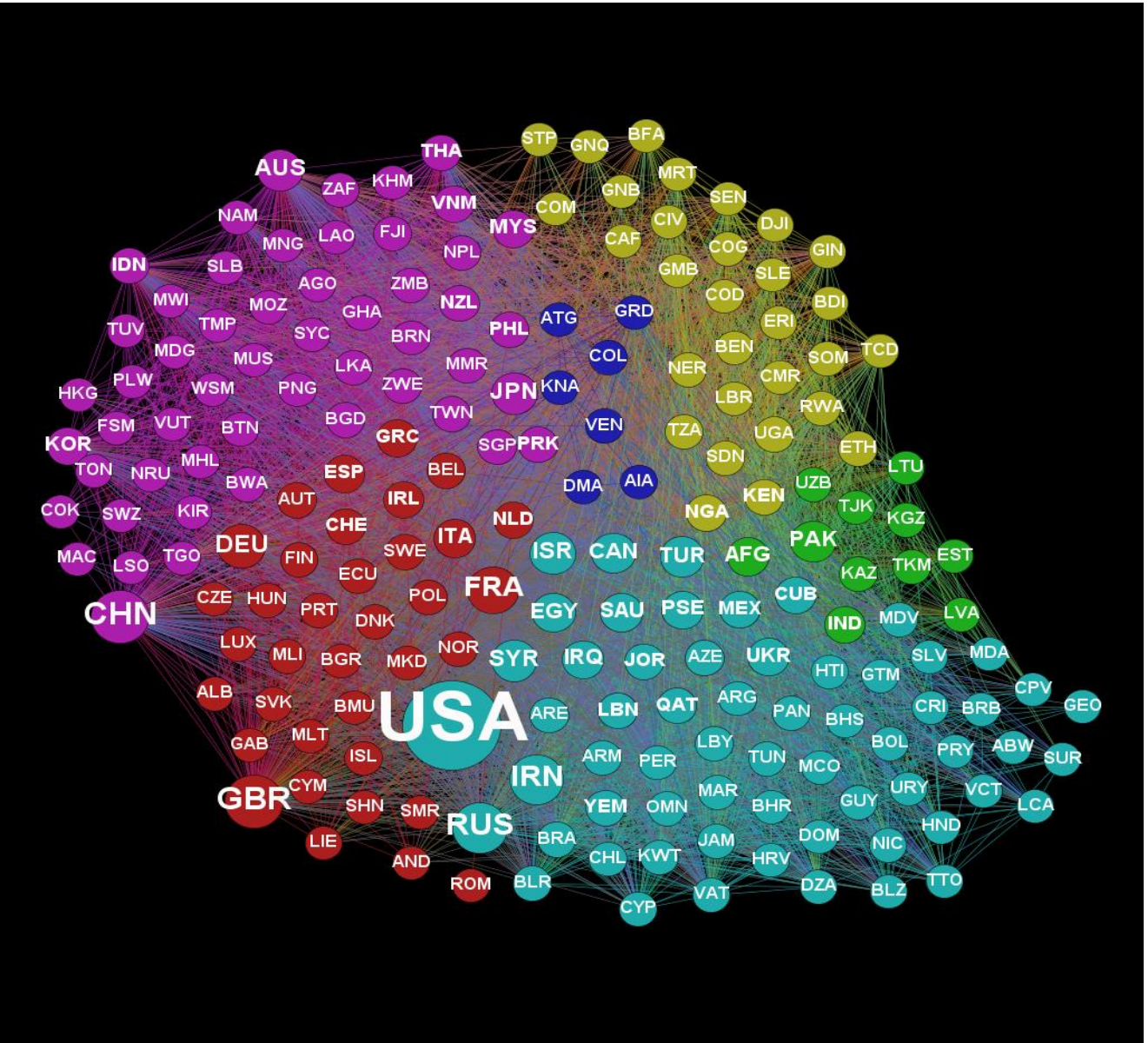
- ✓ To show the variation of the activities for these cures better, we smoothed data by a sophisticated adaptive filter to reduce the resolution of 1 day to that of 3 months.

Tracking the activities in Syria





Clusters detected through global relation matrix



6 communities of the largest principal component computed based on the daily global relationship matrix (G+) from 20150801 to 20150831

Technical challenges

- Need non-negative matrix factorization (Prof. Bertozzi)
- To study dynamical evolution, need to do projection
- Identify major events and follow/examine/predict their evolutions
- Modes of interactions in international relations

  (e.g., making compromise, tip-for-toe, etc; ongoing)
- In politics, prediction is not always possible; the most important is to identify bifurcation-like points
  - Near bifurcation points, different powers have different goals
  - Which one is the most beneficial to the whole mankind?