

# SCIENTOMETRICS

Aparna Basu

[aparnabasu.dr@gmail.com](mailto:aparnabasu.dr@gmail.com)

*<http://sites.google.com/site/aparnabasuhomepage/>*

# Past Affiliations

- National Institute of Science Technology and Development Studies (CSIR) New Delhi  
till Feb 2014
- Institute of Genomics and Integrative Biology (CSIR) New Delhi
- Institute for Defence Studies and Analysis, New Delhi
- IBM Global Services, Bangalore
- NISTADS

# Scientometrics

- Is part of what is called “*the science of science*”
- A body of philosophical, historical, sociological studies on the conduct of science and about scientists.
- Scientometrics is about quantities that can be measured, e.g.,
  - scientific output (*papers, patents*)
  - scientific manpower
  - money invested
- These entities constitute the data that are analysed.

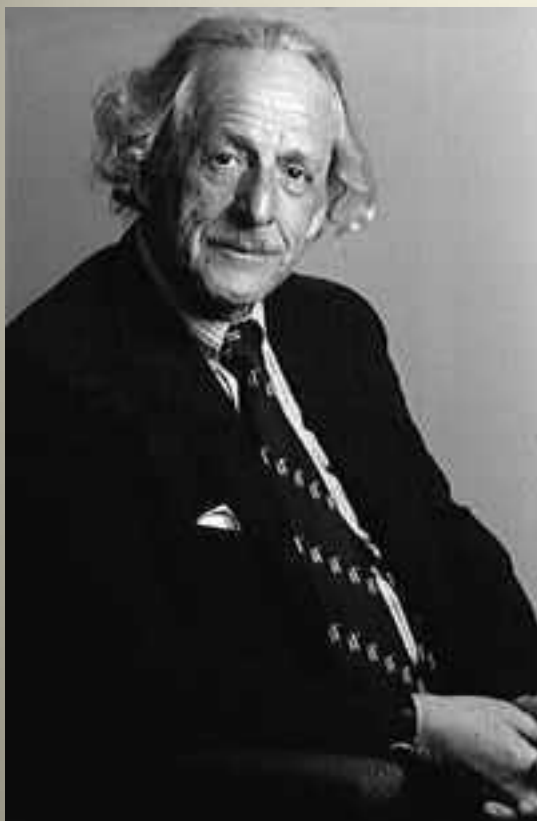
# Early work in Scientometrics/Bibliometrics

- Observation of statistical regularities in the data, e.g.,
  - Power law type relationship in scientific productivity (**Lotka's Law, 1926**)
  - Invariant S-curve observed for the scatter of articles in journals (**Bradford's Law, 1934**)
  - Word frequency in text (**Zipf's Law, 1935,1949** )
  - Distribution of wealth in economics (**Pareto's Law,** )

Basu, A. (1992) Hierarchical Distributions and Bradford's Law, JASIS

# Zipf type distributions-observed contexts

- The sizes of human settlements (few cities, many hamlets/villages)<sup>[10]</sup>
- File size distribution of Internet traffic which uses the TCP protocol (many smaller files, few larger ones)<sup>[10]</sup>
- [Hard disk drive](#) error rates<sup>[11]</sup>
- Clusters of [Bose–Einstein condensate](#) near [absolute zero](#)<sup>[12]</sup>
- The values of [oil reserves](#) in oil fields (a few large fields, many small fields)<sup>[10]</sup>
- The length distribution in jobs assigned supercomputers (a few large ones, many small ones)<sup>[citation needed]</sup>
- The standardized price returns on individual stocks <sup>[10]</sup>
- Fitted cumulative Pareto (Lomax) distribution to maximum one-day rainfalls using [CumFreq](#), see also [distribution fitting](#)
- Sizes of sand particles <sup>[10]</sup>
- Sizes of meteorites
- Numbers of species per genus (There is subjectivity involved: The tendency to divide a genus into two or more increases with the number of species in it)<sup>[citation needed]</sup>
- Areas burnt in forest fires
- Severity of large [casualty](#) losses for certain lines of business such as general liability, commercial auto, and workers compensation.<sup>[13][14]</sup>
- In [hydrology](#) the Pareto distribution is applied to extreme events such as annually maximum one-day rainfalls and river discharges.



## Eugene Garfield

PhD in Structural Linguistics

# science Citation Index

As the President & CEO of the Institute for Scientific Information, Garfield began regular publication of the *Science Citation Index* in 1964.

The *SCI* soon distinguished itself from other literature indexes and was recognized as a basic and fundamental innovation in scientific communication and information science. It was truly current while other traditional indexes were often several years behind the literature.

Although built to aid scientific research, the *SCI* became a tool for scientometrics

# Research Evaluation

## As seen by scientists & managers

- Science and Scientists are the objects of study
- A method of evaluating performance
  - No. of papers, No. of citations, No. of highly cited papers
  - Impact factor of Journals (Nature or Science)
  - Indicators: Citations per paper, Citations per researcher
- Policy
  - National research performance, China vs India

# Institutional Ranking

- **National Ranking: Times, Shanghai, Leiden, Taiwan**
- **ARWU, Times Higher Education, Quacarelli Symonds**
- **Take into account other indicators like teaching, collaboration and internationalization to arrive at a rank**
- **Published once a year or once in two years - only a few Indian institutions ranked in top 500**



# CHANDRASEKHAR S

## Stochastic problems in physics and astronomy

**Author(s):** Chandrasekhar S

**Source:** REVIEWS OF MODERN PHYSICS **Volume:** 15 **Issue:** 1 **Pages:** 0001-0089  
JAN 1943

**Times Cited:** 4,955 **References:** 101 [Citation Map](#) *beta*

**Document Type:** Review

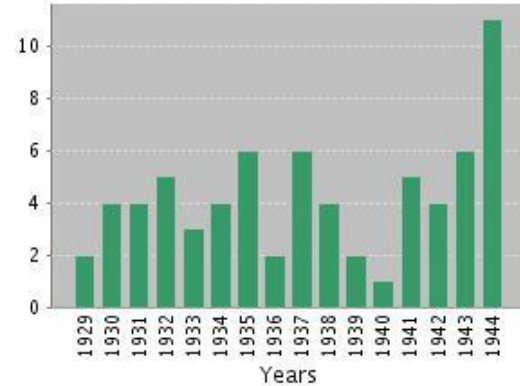
**Language:** English

### Addresses:

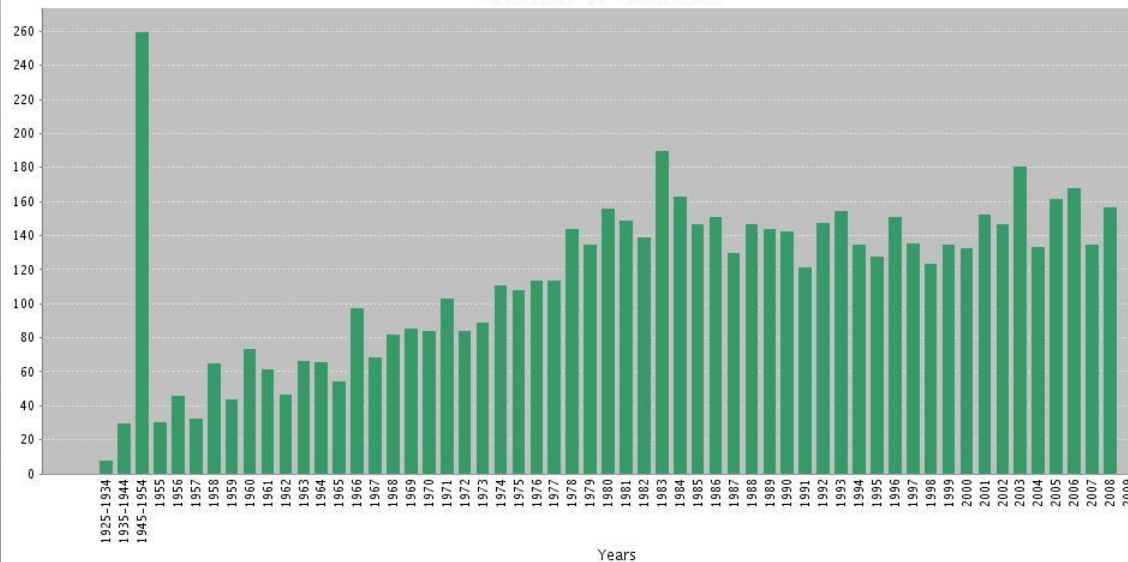
1. Univ Chicago, Yerkes Observ, Williams Bay, WI USA

**Publisher:** AMERICAN PHYSICAL SOC, ONE PHYSICS ELLIPSE, COLLEGE PK, MD 20740-3844 USA

Published Items in Each Year



Citations in Each Year

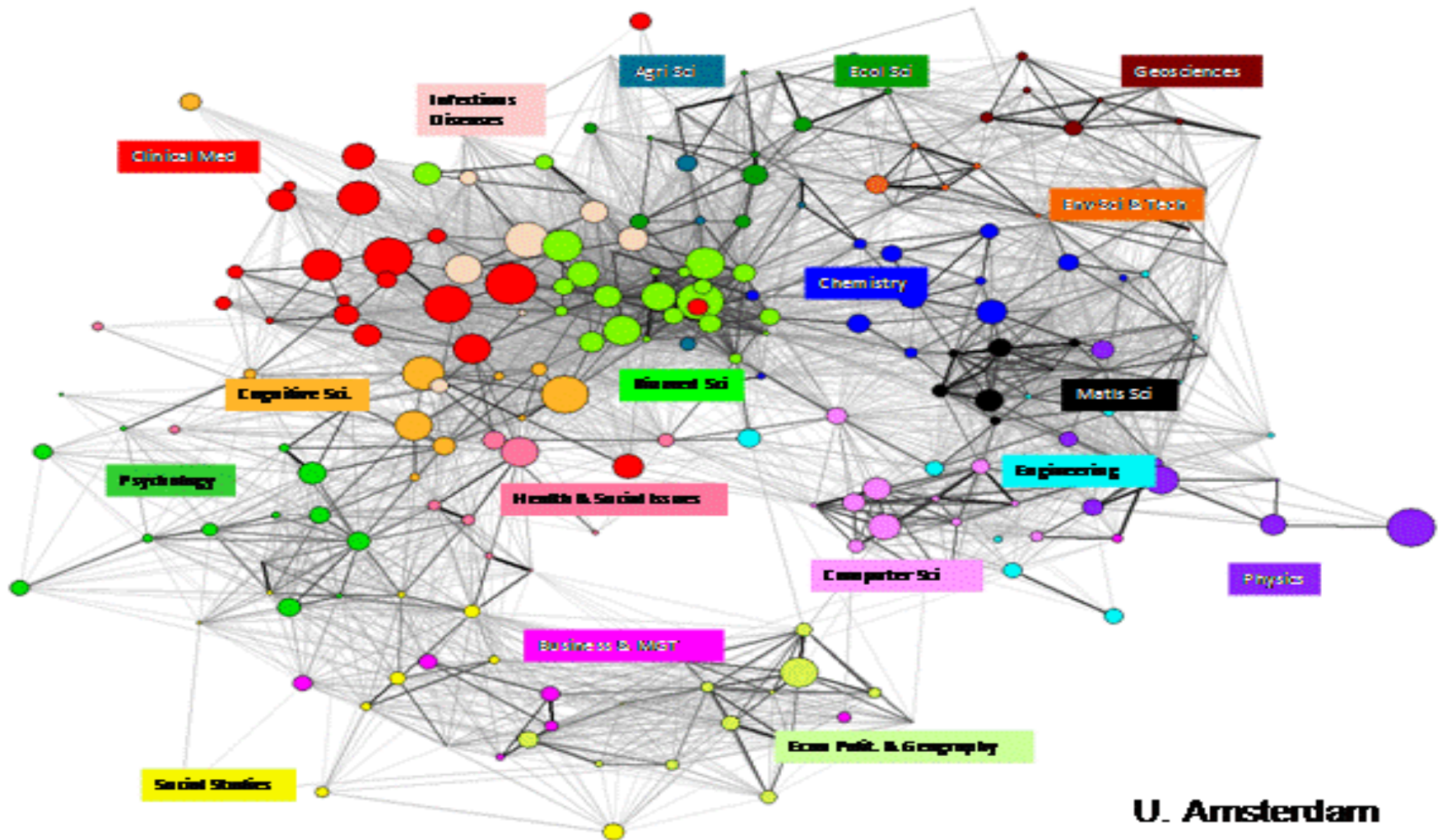


# Network Properties of Scientometric data

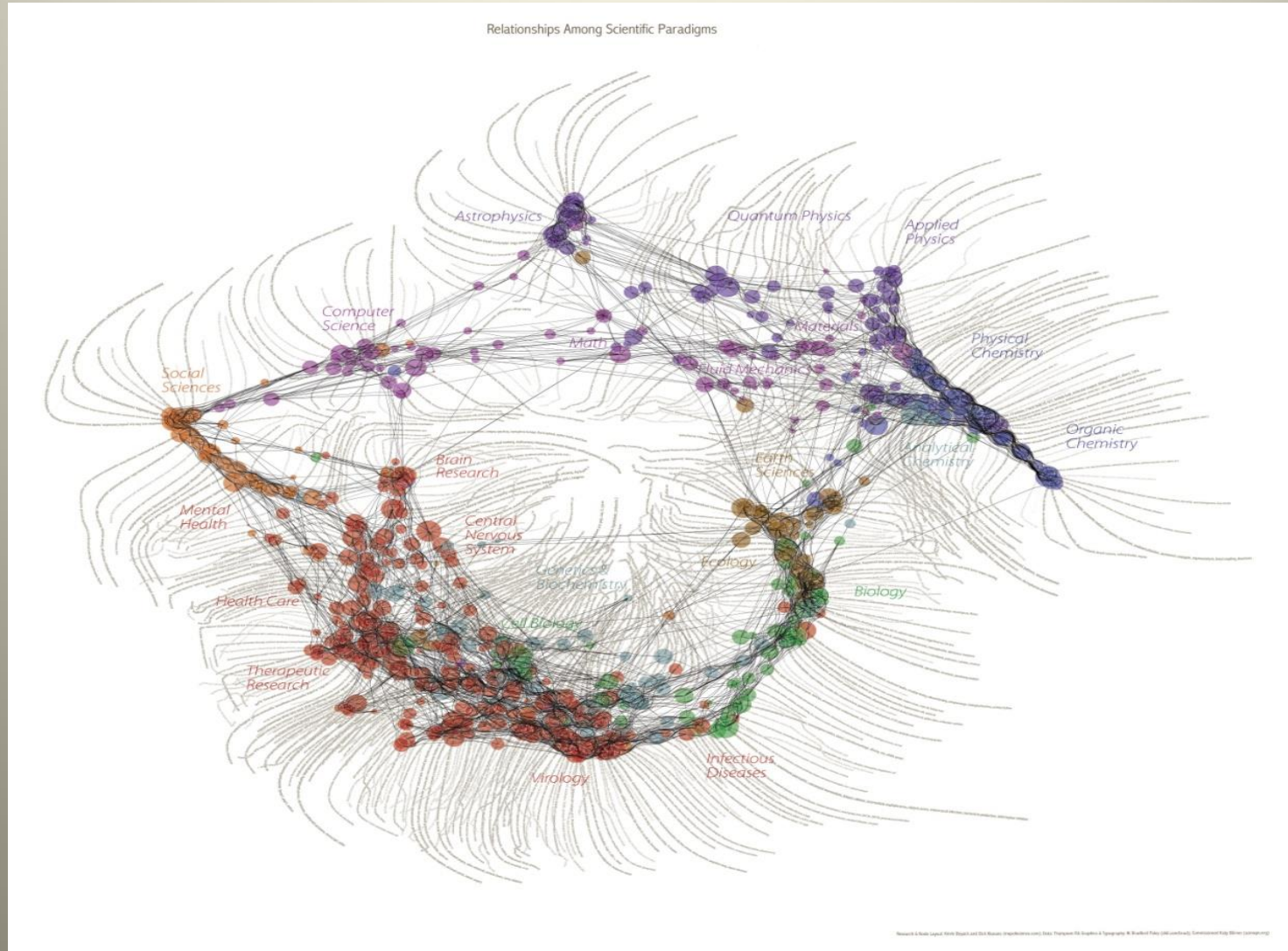
- References or citations
- Collaborative or co-authorship links
- Co-occurrence of words, institutions, countries

Can be computationally intensive

# Network of Science-Boyack & Klavans

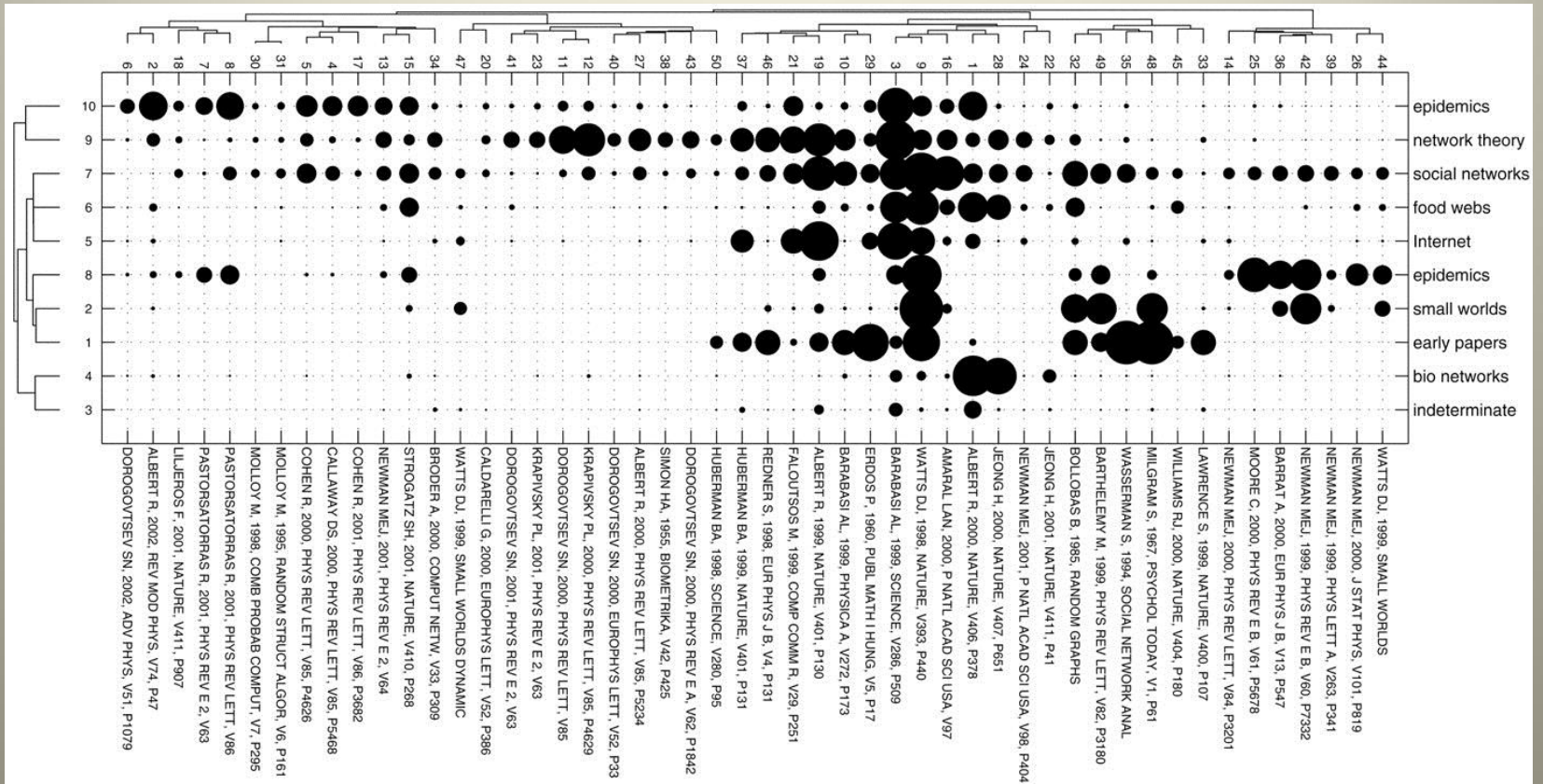


# Trees- Visualization by Michael Sandberg





# Crossmap of research fronts to base reference groups for a collection of complex networks papers.



Steven A. Morris, and Gary G. Yen PNAS 2004;101:5291-5296

## HISTORIOGRAPH OF SMALL WORLD PAPERS 1967-2002

**Nodes: 17, Links: 62**

4 1967 Milgram S, GCS: 148 LCS: 167  
 6 1969 Travers J, GCS: 64 LCS: 43  
 9 1970 Korte C, GCS: 38 LCS: 31  
 194 1998 Watts DJ, GCS: 387 LCS: 134  
 206 1999 Monasson R, GCS: 33 LCS: 20  
 215 1999 Barabasi AL, GCS: 82 LCS: 23  
 218 1999 Newman MEJ, GCS: 74 LCS: 47  
 219 1999 Barthelemy M, GCS: 73 LCS: 47  
 221 1999 Newman MEJ, GCS: 40 LCS: 27  
 224 1999 Barabasi AL, GCS: 269 LCS: 49  
 229 2000 Barrat A, GCS: 58 LCS: 39  
 243 2000 Newman MEJ, GCS: 39 LCS: 23  
 249 2000 Albert R, GCS: 154 LCS: 20  
 257 2000 Moore C, GCS: 48 LCS: 25  
 263 2000 Newman MEJ, GCS: 35 LCS: 26  
 265 2000 Amaral LAN, GCS: 111 LCS: 28  
 300 2001 Strogatz SH. GCS: 128 LCS: 23

1967 (4)



1969 (2)



1970 (4)



1998 (15)



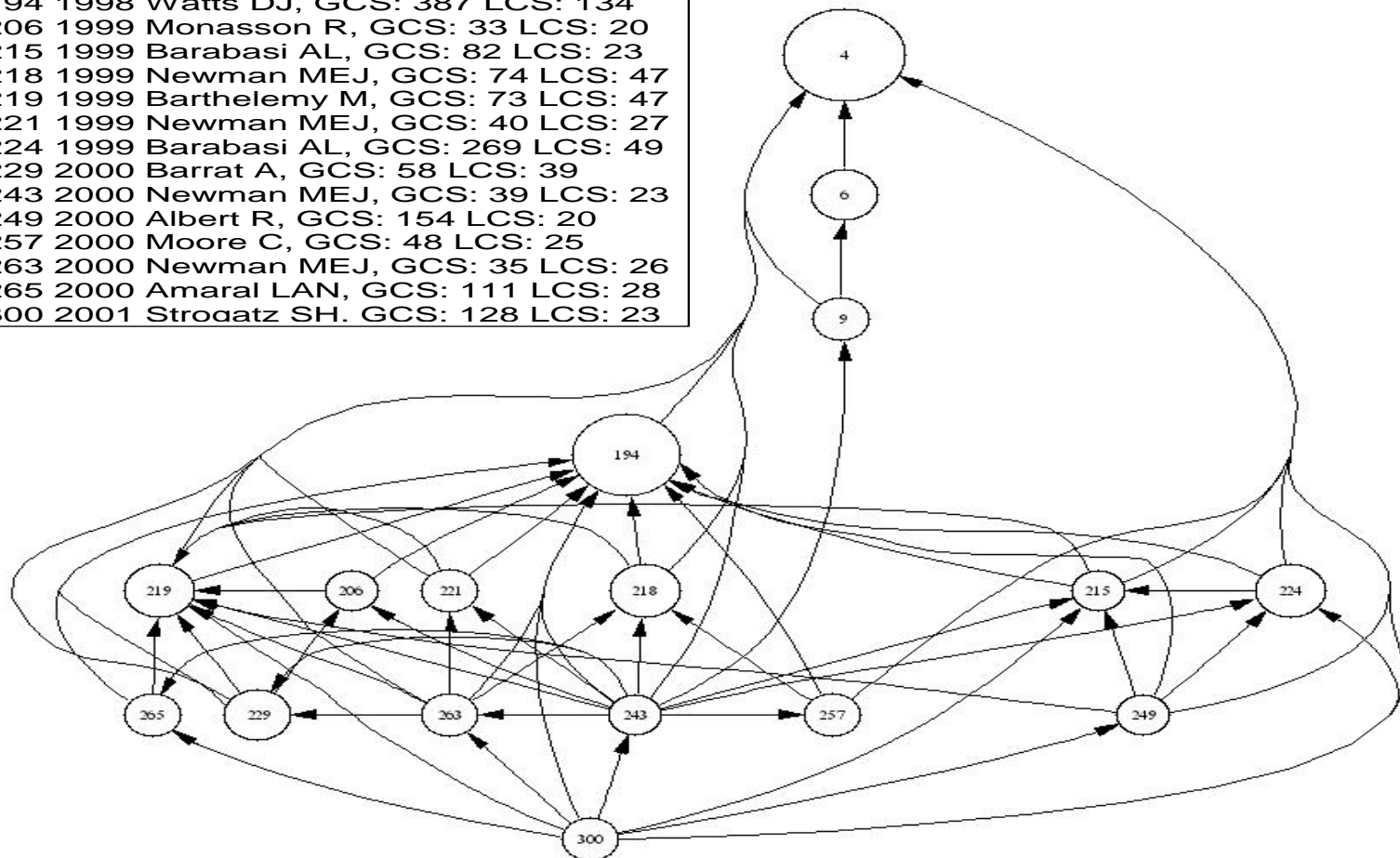
1999 (28)



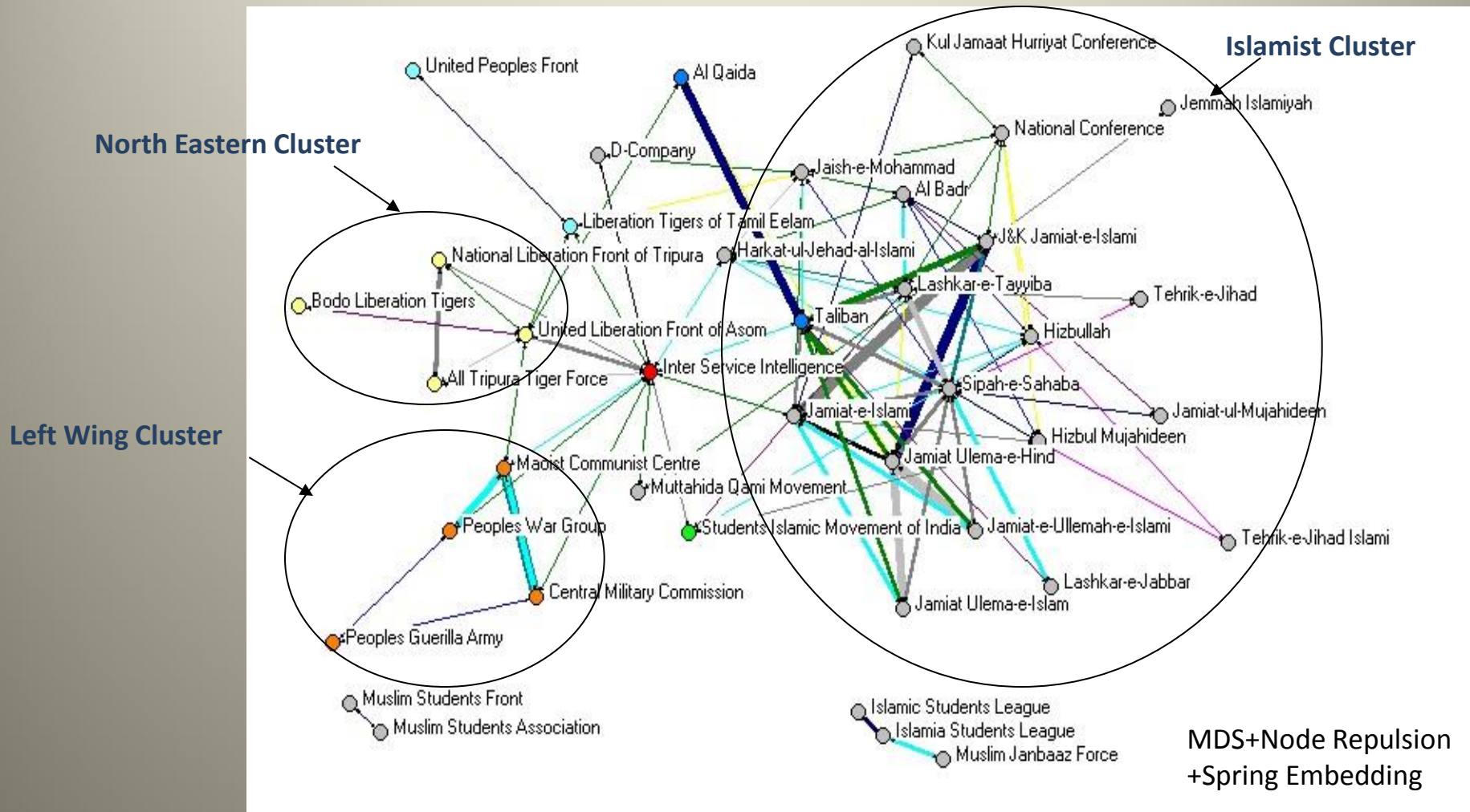
2000 (46)



2001 (58)

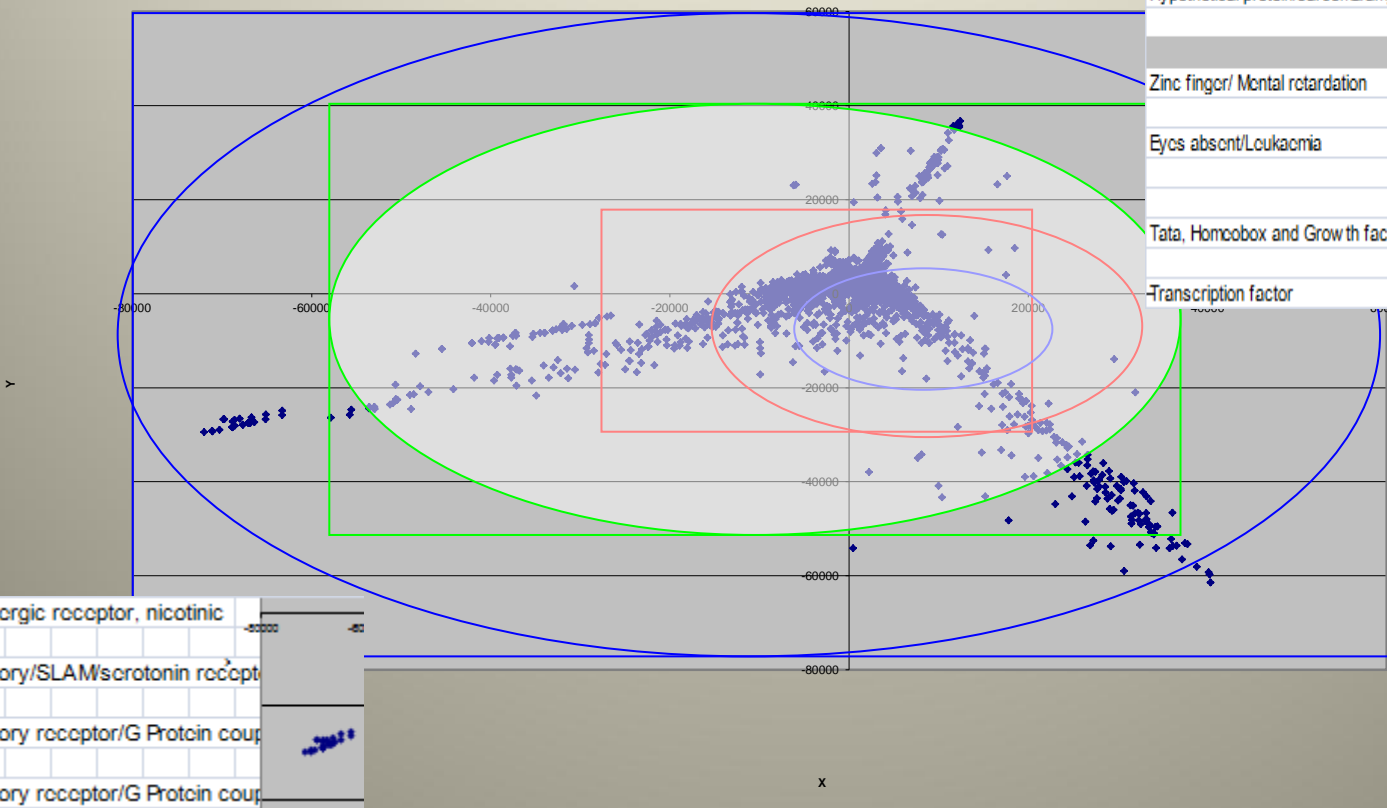


# Ethnic and ideological clusters of terrorist organizations in India



# Schizophrenia Candidate Genes MDS

MDS: Schizophrenia genes



Hypothetical protein
Hypothetical protein/sarcoma/amyotrophic lateral sclerosis
Zinc finger/ Mental retardation
Eyes absent/Leukemia
Tata, Homeobox and Growth factors
Transcription factor

Cholinergic receptor, nicotinic
Olfactory/SLAM/serotonin receptor
Olfactory receptor/G Protein coupled
Olfactory receptor/G Protein coupled



# Thank you

Contact

[aparnabasu.dr@gmail.com](mailto:aparnabasu.dr@gmail.com)

*<http://sites.google.com/site/aparnabasuhompage/>*