



# Big Data meets Network Science

SAKIB MATIN

DECEMBER 2, 2016

# Overview of Presentation

- ▶ Motivation
- ▶ Back Ground
- ▶ Source of Data
- ▶ Results from Econophysics
- ▶ Results from Network Science
- ▶ Conclusions
- ▶ Future Works

# Motivation

Many behaviors and patterns were undiscovered because of small data sets and limited computational powers. Cheap storage and abundant computational power has motivated data driven approach to problem solving.

1996 Nature paper marked beginning of data driven Econophysics. The findings of the paper challenged certain assumptions present in economic theories.

# Motivation



Network Science allows us to map real world problems onto networks and use powerful algorithms to solve them.

Tools created by Network Science, allows us to easily visualize complex structures and summarize key information about a system.

# Background

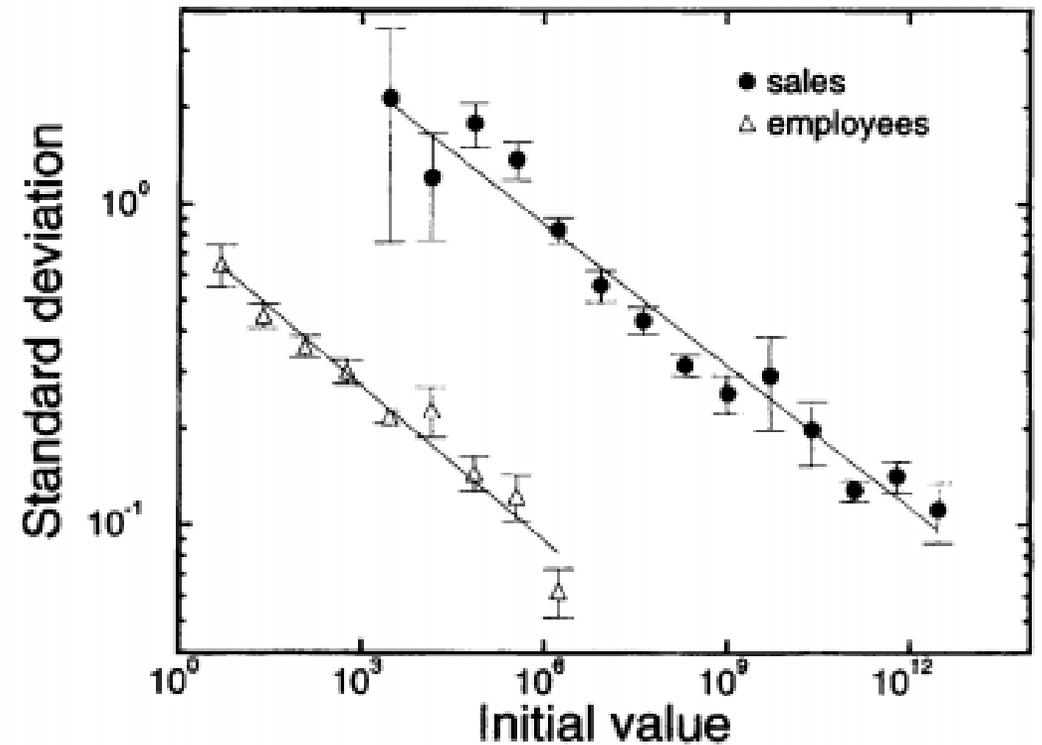
“Scaling behavior in growth of companies” was published in Nature in 1996. It is a phenomenological paper that looks at growth of publicly traded firms. The data driven approach revealed many complex patterns and connections in economics.

Firm growth was shown to obey certain scaling laws over many orders of magnitude. The results are highly robust and point towards a universal theory.

# Back Ground

Standard-Deviation of 1-year growth rates as a function of initial value.

Slopes are  $\sim 0.16$

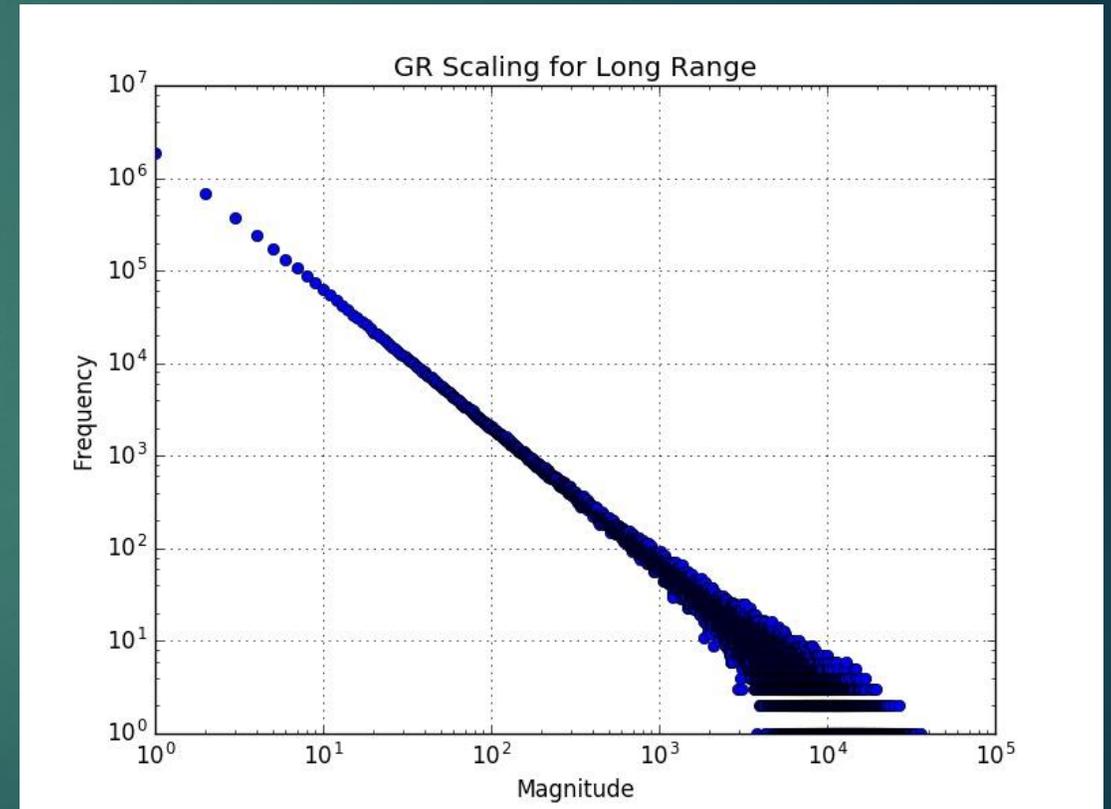


# Power Laws

The functional form of power laws is:  $f(x) = a x^{-k}$

An important attribute of power laws is scale invariance. Physical systems near a critical point are characterized by scale invariance.

On a log-log scale, a power law is plotted as a straight line.



# Power Laws in Nature

In physics, power laws are ubiquitous in the study of critical phenomena. Many thermodynamic response variables diverge as a power law near the critical point.

Power laws can also be found in other fields. Zipf's law is a power-law relationship between rank and magnitude of characteristic variable (such as frequency, size etc.)

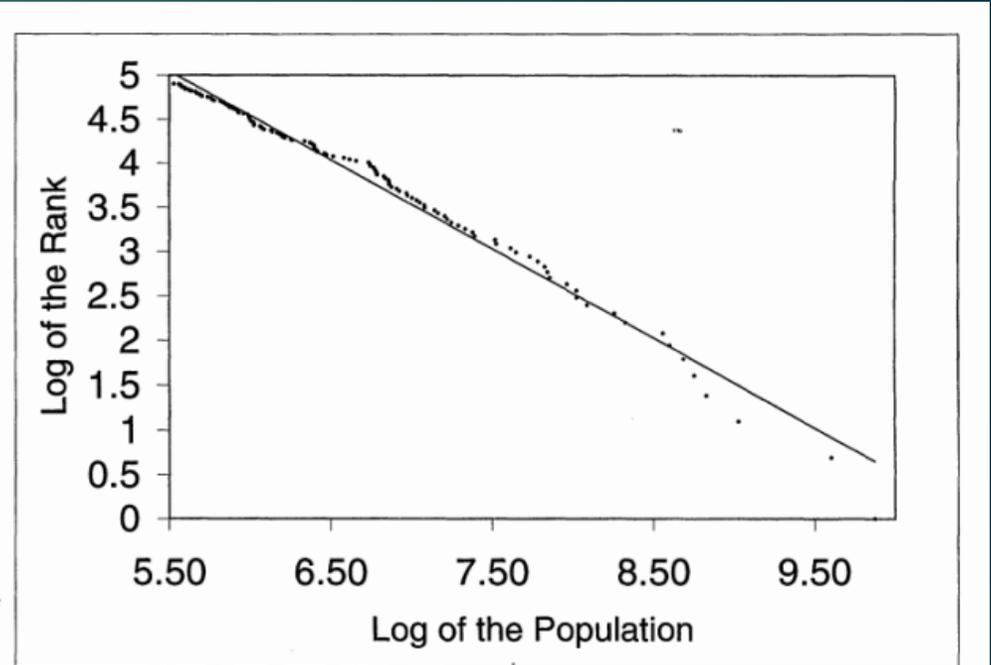


FIGURE I  
Log Size versus Log Rank of the 135 largest U. S. Metropolitan Areas in 1991  
Source: Statistical Abstract of the United States [1993].

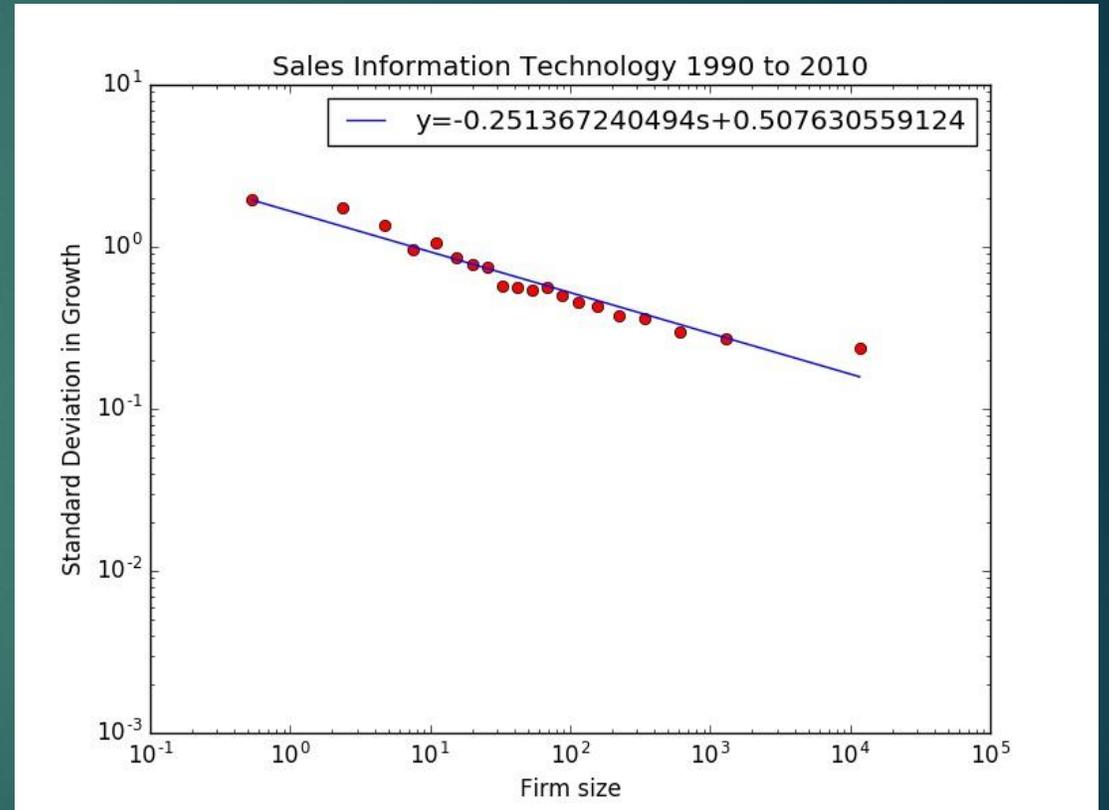
# Source of Data

**Compustat** began in 1961 and is a database of financial, statistical and market information on active and inactive global companies throughout the world.

The database covers 99,000 global securities, covering 99% of the world's total market capitalization with annual company data history available back to 1950.

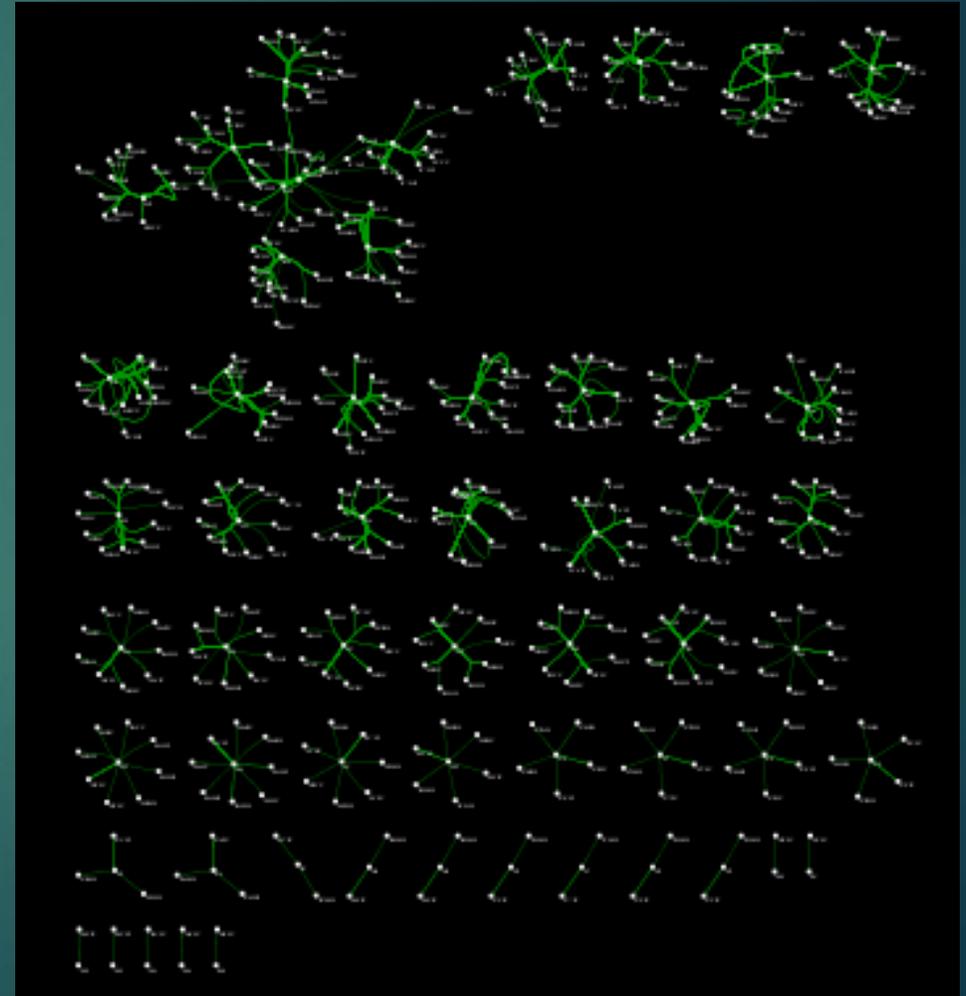
# Scaling Law in Economics

- ▶ Good straight-line fit.
- ▶ Some fluctuations at end-points to smaller number of data points.
- ▶ Result consistent over 5 orders of magnitude!



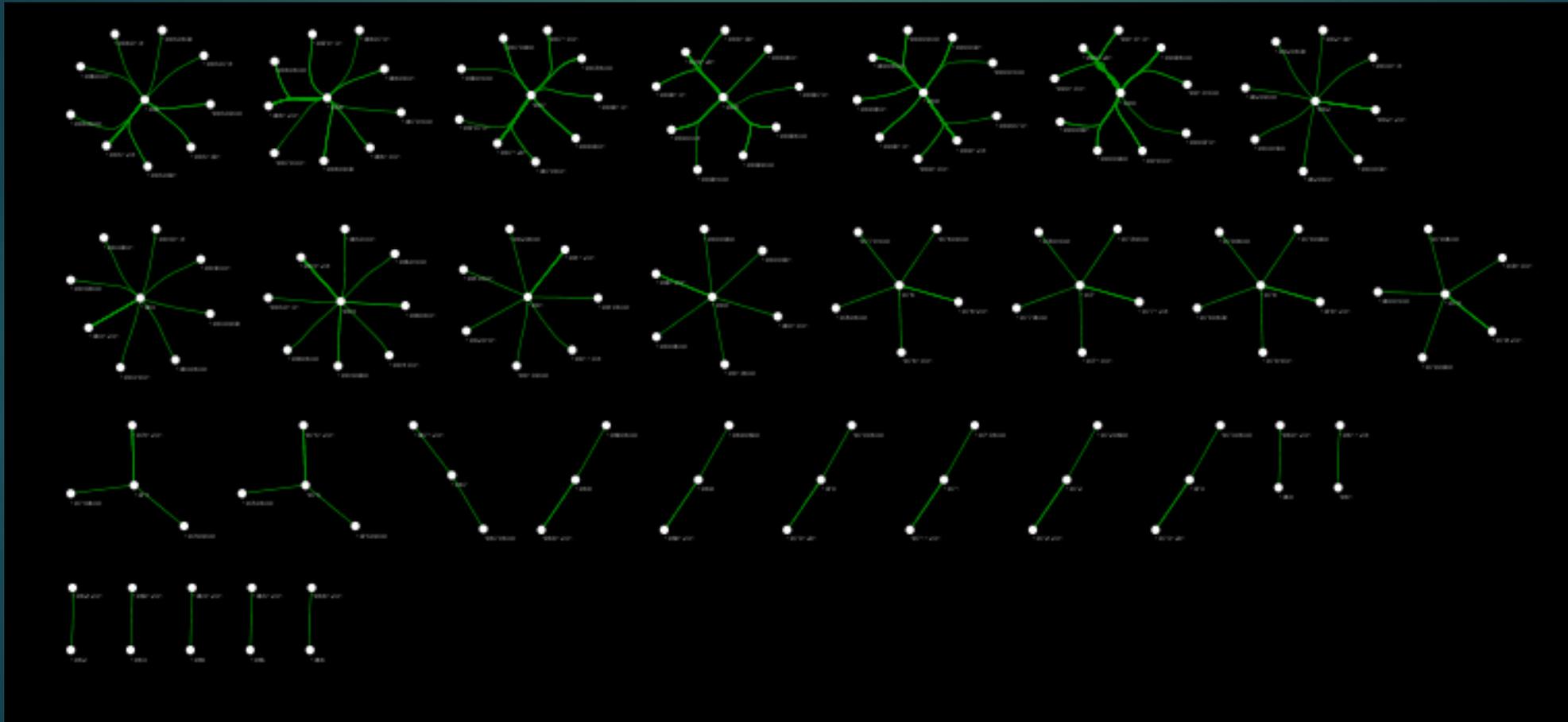
# Case Study

- ▶ 46 years represented.
- ▶ Node represents data collection source.
- ▶ Increasing complexity.
- ▶ Recent years data has to pass through intermediate agencies.
- ▶ Reflects diversifying of the IT sector.



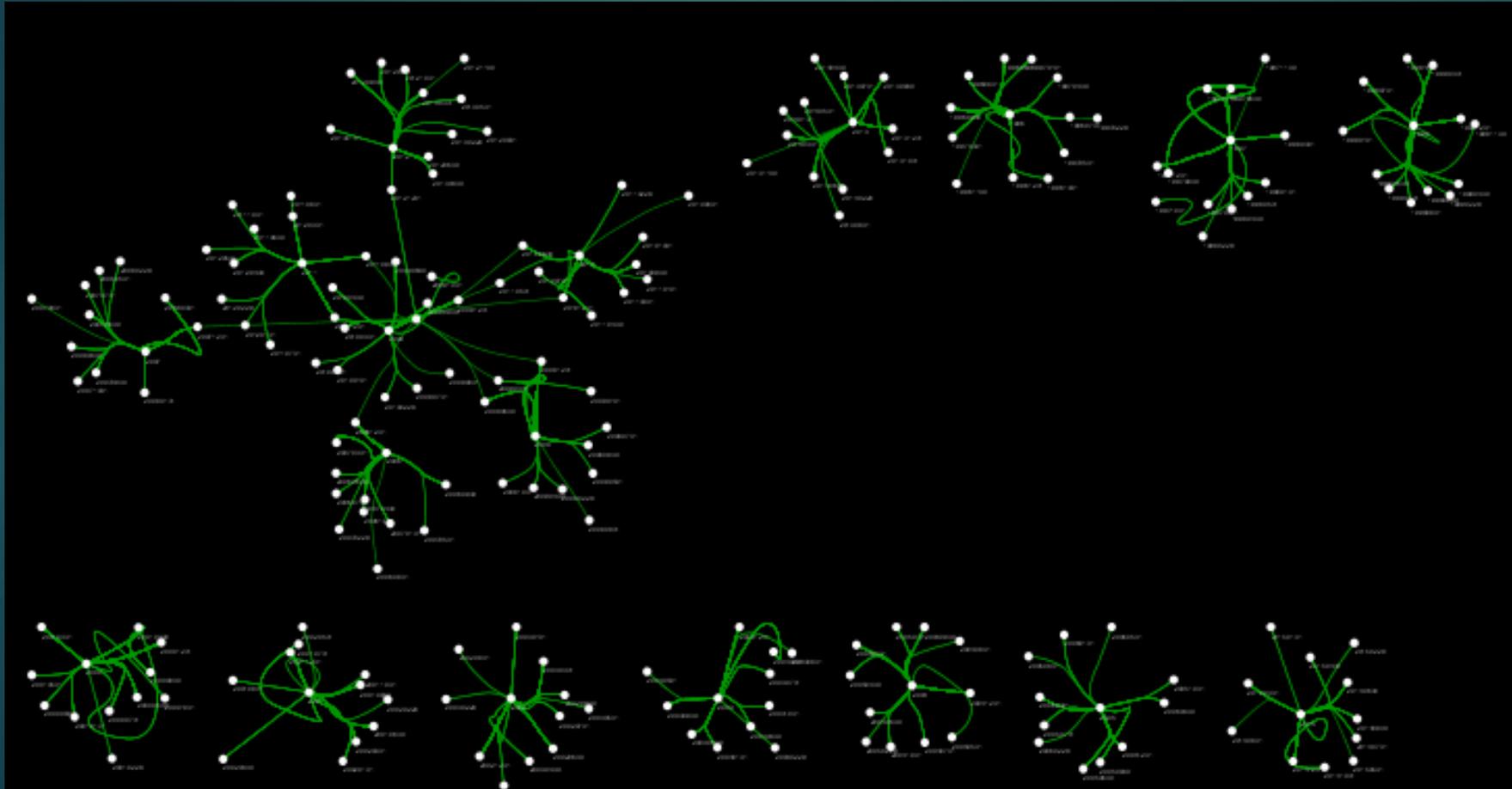
# Case Study

## Beginning of IT industry



# Case Study

## Recent Years



# Conclusions

- ▶ Scaling Law in Economics.
- ▶ Organization of Complex behavior in the market
- ▶ Time- evolution of Market structure

# Open Questions

- ▶ External factors which contribute to 'scaling'.
- ▶ A more systematic analysis for world wide companies and stock markets
- ▶ Dynamical Evolution of Networks: Are there Universal trends? Effects of Regulation?