Creen Meeting in Karlsruhe, June 2004

Agata Fronczak, Piotr Fronczak and Janusz A. Hołyst

Faculty of Physics, Warsaw University of Technology

Information-theoretic approach to random networks: fluctuations and responses

CREEN - Critical Events in Evolving Networks

Part B1 of the proposal: Scientific and technological objectives of the project

The objective of this project is:

a) to develop new methods to recognize emerging critical events in complex networks

b) to apply these methods to the analysis of the emergence of new research topics (scientific avalanches) and possible crises in a social institution – the public trust in science

From statistical physics we know that

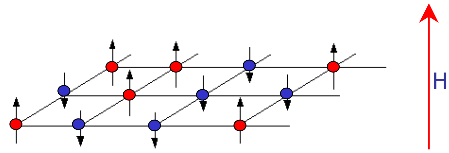
a system entering into the critical region exhibits large scale fluctuations

and is very sensitive to presence of external fields.

T1.1. *Description of static and dynamic correlations functions for network structure and nodes internal dynamics*. We will check if special kinds of fluctuation-dissipation theorems exist that join correlation functions with network susceptibility to external forces and allow to predict network response in the face of environmental perturbations. (...)

Fluctuation-Dissipation Relation for Magnetization in Ising Model

The Ising Model:



Applications
- ✓ simple magnet
- ✓ opinion formation model;

Definition of the model:
- ✓ Suppose one have a network (lattice)
- ✓ Each site in such a network can have two states ↑ or ↓ ($s_i = \pm 1$)
- ✓ Neighboring nodes have an energetic preference to have the same value
- ✓ External force H that supports one state (↑ or ↓)

$$M = \sum_{i=1}^{N} s_i$$

Order parameter – which measures how many nodes have the same state

$$\chi = \frac{\partial \langle M \rangle}{\partial H}$$

Magnetic susceptibility measures the strength of the response of the magnetization *M* to changes in the field H

Fluctuation–Dissipation Relation

$$\chi = \frac{\partial \langle M \rangle}{\partial H} = \frac{\langle M^2 \rangle - \langle M \rangle^2}{kT}$$

Fluctuation-Dissipation Relation ...

$$\chi = \frac{\partial \langle M \rangle}{\partial H} = \frac{\langle M^2 \rangle - \langle M \rangle^2}{kT}$$

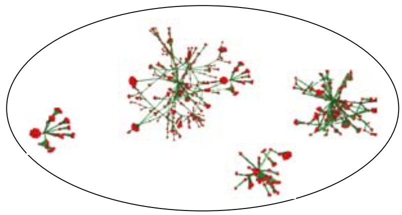
The fluctuation-dissipation theorems are interesting for a number of reasons:

1. They join both microscopic description and macroscopic properties of the considered system.
2. They relate the actual state of the system (fluctuations) to its future behavior (response).
3. Due to fluctuation-dissipation relations phase transitions certified by singularities in susceptibilities can also be reported by large scale fluctuations.

It is not a trivial task how to perform averages over random networks

Random networks as random representatives of an ensemble

The same meaning as in the case of a random variable



- **Specify a set of graphs** which one wants to study (e.g. simple graphs, digraphs, weighted graphs)

$$\Omega = \{G_1, G_2, \ldots\}$$

- **Establish probability distribution** $P(G)$ over the ensemble.

Information-theoretic approach to random networks (1)

Up to now, there has been proposed different weight assignment strategies to networks (mainly due to Z. Burda, S. Dorogovtsev, I. Farkas and M.E.J. Newman)

In our study we take advantage of the maximum-entropy principle.

We maximize the Shannon entropy

$$S = -\sum_G P(G) \ln P(G),$$

subject to the constraints

$$\langle x_i \rangle = \sum_G x_i(G) P(G)$$

for $i = 1, 2, \ldots, r$, plus the normalization condition

$$\sum_G P(G) = 1. \qquad (3)$$

Constraints

1. *Trivial*: Expected number of connections $\langle m \rangle$
2. *Realistic*: Expected degree sequence $\{\langle k_1 \rangle, \langle k_2 \rangle, \ldots, \langle k_N \rangle\}$ (e.g. resulting in SF degree distribution)

Information-theoretic approach to random networks ...

The Langrangian for the above problem is given by the below expression

$$\mathcal{L} = -\sum_G P(G)\ln P(G) + \alpha(1 - \sum_G P(G)) + \sum_{i=1}^{r}\theta_i(\langle x_i\rangle - \sum_G x_i(G)P(G)), \qquad (4)$$

where the multipliers α and θ_i are to be determined by (2) and (3).

Differentiating $\mathcal{L}$ with respect to $P(G)$ and then equating the result to zero one obtains the desired probability distribution over the ensemble of graphs with given properties (2)

$$P(G) = \frac{e^{-H(G)}}{Z}, \qquad (5)$$

where $H(G)$ is the network Hamiltonian

$$H(G) = \sum_{i=1}^{r}\theta_i x_i(G), \qquad (6)$$

and Z represents the partition function

$$Z = \sum_G e^{-H(G)} = e^{\alpha+1}. \qquad (7)$$

Information-theoretic approach to random networks (3)

Maximum-entropy networks with an expected degree sequence

$$\{\langle k_1\rangle, \langle k_2\rangle, \ldots, \langle k_N\rangle\}$$

In this case the network Hamiltonian is

$$H(G) = \sum_{i=1}^{N}\theta_i k_i(G),$$

where

- $k_i(G)$ represents degree of the node i
- θ_i is conjugate field that is coupled to degree.

(like magnetization M and external field H)

Network of scientists:

What is the field which couples to the connectivity of a scientist?

age, # publications, # citations, # patents, money

One can prove the following fluctuation-dissipation relations:

1.

$$\chi = \frac{\partial\langle k_i\rangle}{\partial\theta_i} = \langle k_i^2\rangle - \langle k_i\rangle^2$$

where the averages are taken over all nodes that have the same value of the assigned potential (hidden attribute).

2.

$$\chi_{ij} = \frac{\partial\langle k_i\rangle}{\partial\theta_j} = \langle k_i k_j\rangle - \langle k_i\rangle\langle k_j\rangle$$

where the nodes k_i are assumed to have the same potential h_i, whereas the nodes k_j are characterized by h_j.

A few ideas

$$H(G) = \sum_{i=1}^{N} \theta_i k_i(G)$$

$$\chi = \frac{\partial \langle k_i \rangle}{\partial \theta_i} = \langle k_i^2 \rangle - \langle k_i \rangle^2$$

Population of scientists:

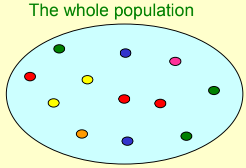
Points – scientists

Parameters characterizing scientists – # papers (k_i)

Field (i.e. hidden attribute) coupled to the # of papers – year of the first publication (theta_I) - (i-index of a subpopulation)

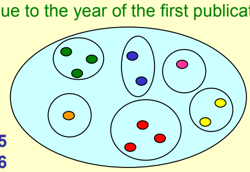
(Is the 'year of the first publication' the correct field coupled to the #papers???)

The whole population



Field:
1975
1976

Specification of sub-populations due to the year of the first publication



Left-H-S: ***differential derivative***

Right-H-S: ***variance***

A few ideas

$$H(G) = \sum_{i=1}^{N} \theta_i k_i(G)$$

$$\chi = \frac{\partial \langle k_i \rangle}{\partial \theta_i} = \langle k_i^2 \rangle - \langle k_i \rangle^2$$

Network of scientists:

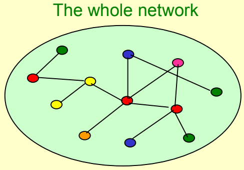
Nodes – scientists

Parameter characterizing a scientist – # collaborators (degree k_i)

Field (i.e. hidden attribute) coupled to the # collaborators – # citations (theta_I) - (i-index of a subpopulation)

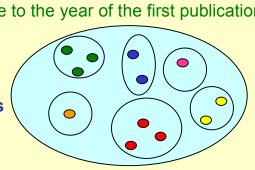
(Is # citations the correct field coupled to the #collaborators ???)

The whole network



#citations
0-20
20-50

Specification of sub-populations due to the year of the first publication



Left-H-S: ***differential derivative***

Right-H-S: ***variance***