Extreme Edges of the Internet

Ingve Simonsen
Department of Physics
NTNU, Trondheim, NORWAY

Work done in collaboration with:
Kasper A. Eriksen, Lund
Sergei Maslov, Brookhaven
Kim Sneppen, Nordita/NBI
Motivation

• For many real-world networks it is important to:
  – have information on their modular structure
  – know how to increase their stability and robustness
  – identify critical links

• This is what we set out to do with diffusion!

• Physical Motivation:
  – Consider diffusion on complex networks (ensemble of Random Walkers)
  – Steady state will be reached first \textit{locally}, then \textit{globally}

• Practical example:
  – How easy is it by \textit{random driving}, to reach South America if you start in North America?
Outline of the talk

• Background Theory
  – Diffusion Equation for Complex Networks
  – Participation Ratio
• Real-World Examples
  – Zachary’s friendship network
  – Internet network (Autonomous System)
• Application
• Conclusions
Diffusion on Complex Networks

• Conservation of walker density:

\[ \rho_i(t+1) - \rho_i(t) = \sum_j A_{ij} \frac{\rho_j(t)}{k_j} - \sum_j A_{ji} \frac{\rho_i(t)}{k_i} \]

• In matrix form (\(T\) : transfer matrix, \(k\): connectivity of node \(i\))

\[ \rho(t+1) = T \rho(t) \quad T_{ij} = \frac{A_{ij}}{k_j} \]

• A steady state exists, and is given by the connectivities:

\[ \rho(t) = T^t \rho(0) \Rightarrow \rho_i(\infty) \propto k_i \]
Diffusion on Complex Networks, \textit{cont.}

- The Transfer matrix is \textit{not symmetric}, but it is similar to the symmetric matrix

\[ S = KTK^{-1} \]

\[ K_{ij} = \frac{\delta_{ij}}{\sqrt{k_i}} \]

- Since $S$ is symmetric, the transfer matrix has
  - real eigenvalues, $\lambda^{(\alpha)}$, (sorted in decreasing order)
  - due to conservation of walkers $|\lambda^{(\alpha)}| \leq 1$
  - and $\lambda^{(1)} = 1$ corresponds to the \textit{stationary state}

- \textit{Outgoing} currents flowing from node $i$ along each of its links
  (Technically, $c^{(\alpha)}$ is the eigenvector for $T^T$):

\[ c_i^{(\alpha)} \propto \frac{\rho_i^{(\alpha)}}{k_i} \]

\[ \text{Stationary state : } c_i^{(1)} = \frac{1}{\sqrt{N}} \]
Participation Ratio

• For a normalized (link) current, \( \sum_i (c_i^{(\alpha)})^2 = 1 \), the participation ratio (PR) is defined as:

\[
\chi_\alpha = \left[ \sum_{i=1}^{N} (c_i^{(\alpha)})^4 \right]^{-1}
\]

• For the steady state: \( \chi_1 = N \)
• The PR gives a measure of the number of nodes participating significantly in a given eigenmode
• If the mode is related to a module, PR will roughly measure its size
The Null Model: A Randomized Network

• **Important**: The randomization process should **not** alter the *degree distribution* of the network!

• **The algorithm**:
  - Choose two arbitrary nodes (A and C)
  - Choose one of their nearest-neighbors (or links) randomly (B and D)
  - **Rewire**: $A \rightarrow D$ and $C \rightarrow B$ if these links do not exist

Results for Real-World Networks

Two examples:

- Zachary’s Karate club Network (friendship network)
  - The university karate club breaks up due to an internal conflict
  - A small network (N=34; L=72)
  - *The modular structure is known!*

- Course-grained Internet Network (Autonomous Systems)
  - Medium sized network (N=6,474; L=12,572)
  - *The modular structure is not well-known in advance*
  - Reference: National Laboratory of Applied Network Research
    http://moat.nlanr.net/AS/
Zachary’s friendship Network

These are the assignment made by Zachary!

We use unweighthed links!

N=34

[From PNAS 99, 7821 (2002)]

[Diagram showing the network with N=34 nodes and connections labeled as Trainer w/supporters and Administrator w/followers]

Ingve Simonsen

May 04
Zachary’s friendship Network, cont.

[From PNAS 99, 7821 (2002)]

Node 3 is not correctly identified!
Weighted Zachary network
Summary: Zachary’s friendship Network

The known modular structure is well reproduced by the diffusion model!

Measured Cluster sizes:

\[ \chi_2 = 14 \quad 0 \]
\[ \chi_3 = 13 \quad 3 \]

Theoretical sizes:

\[ N_1 = 18 \]
\[ N_2 = 16 \]
We now consider a larger network!

\[ N = 6,474 \]
\[ L = 12,572 \]

**Definition:** An Autonomous System (AS) is a connected segment of a network consisting of a collection of subnetworks interconnected by a set of routers.

**Nodes:** AS numbers

**Links:** Information sharing
The Internet is indeed modular.

The modules correspond rough to the national structure.

The extreme edges of the Internet are represented by Russian and US military sites!

These country modules could not have been detected using spectral analysis of A

Ref: PRE 64, 026704 (2001)
Autonomous System Network, cont.

- A Module is defined by a PR much larger than its random counterpart (RSFN)
- Number of Modules $M=100$
- Modularity

$$\sum_{\alpha=1}^{M} \chi_{\alpha} = 5400$$

and avoiding double counting gives:

$$\frac{1800}{N} = 30\%$$

**Binning** $\Delta \lambda = 0.05$
Applications

the way to search the WED..

The advantage of Google:

*Popular web pages are listed first among the search results!*

- This is achieved by associating a PageRatio with each page related to the number of links pointing to this page.

- In our diffusion language this means

\[
\text{PageRatio}(i) = \rho_i(\infty)
\]

Conclusions

- Generalized diffusion to (discreet) complex networks
- Diffusion goes beyond next-nearest neighbor interaction

- Shown that diffusion may assist in probing the network’s (large scale) topology
- Participation ratio is an important quantity that contains information about:
  - the number of significant modules (relative to a randomized network)
  - the size of the modules
The End

Reference:

K. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen

*Modularity and extreme edges of the Internet*