Combinatorial Models of Complex Systems

M.G. Beiró

Facultad de Ingeniería
Universidad de Buenos Aires

November 7th, 2013

Complex Networks and Data Communications Group

Topics

Topics

Complexity Sciences

Complexity

Large number of variables -

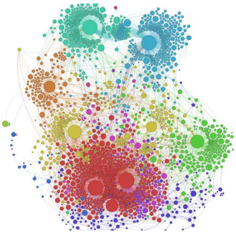
Unformalized relations -

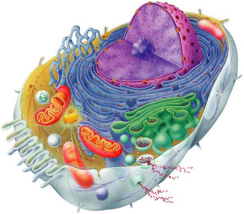
Organic (macroscopic) description -

Interdisciplinarity -

What makes a system *complex*?

- Emergence
  - Collective behavior, global function
  - Not present in the individual components
- Self-Organization
  - Presence of an ordered structure
  - Spontaneous and decentralized





Example: Social Networks

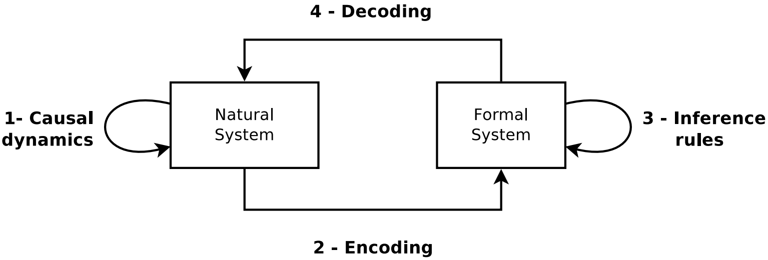
Properties

- Modular structure
  - Communities
- Small-world phenomenon
  - Small average distances
  - High clustering
- Assortative mixing

Relevant problems

- Spreading (information, epidemics, rumors)
- Predict group behavior (segregation, mixing, opinion)
- Understand social relationships

Modeling complex systems



Robert Rosen's modeling relation

Aims of this work

We aimed at...

- Studying combinatorial models
- Proposing efficient models
- Evaluating the performance
- Evaluating scalability
- Visualizing results

Topics

Our contributions

1. Connectivity in the Internet
2. Clustering in networks
3. Social networks: Discovery of community structure

Internet

Two main levels

- The inter-router level (IR)
- The Autonomous Systems level (AS)
  - Internet backbone

    e.g., LEVEL3, HURRICANE, ...
  - Transit ASes (e.g., medium-sized ISPs)
  - Stubs

Internet

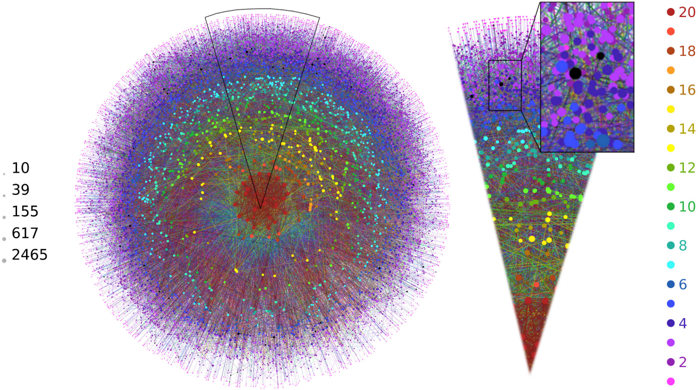
Some results on the Internet topology

- Design of exploration techniques
  - CAIDA, DIMES, RouteViews projects
- Power-laws on the degree distribution
  - Both at the IR-level and the AS-level [FFF99]
- Disassortative mixing
  - Disassortative behavior at the AS-level [PSVV01]
  - Robust-yet-fragile
- Design of hierarchical models
  - Jellyfish model [STF06]
  - MEDUSA model [CHKS07]
  - k-core decomposition [AHDBA08]

Revealed a rich hierarchical structure

Internet

Our contribution: Estimating Internet AS connectivity

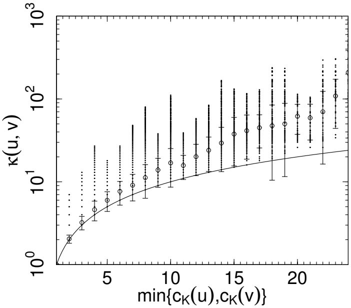


(picture generated with LaNet-vi)

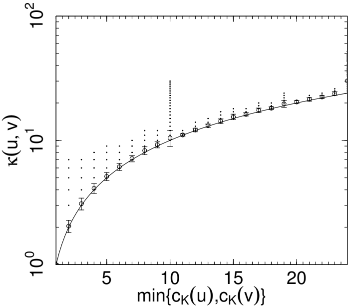
- Sufficient conditions for assuring connectivity k in k-cores [AHBB11]

Internet

Our contribution: Estimating Internet AS connectivity



Total connectivity



Connectivity through core

$O(n(G)e(G))$ vs. $O(e(G))$

Clustering

Clustering:

- Presence of triangles in the network
- Consequence of correlations between nodes
- Effect on the structural and dynamical properties

The Watts-Strogatz model was the first with high clustering coefficients [WS98]

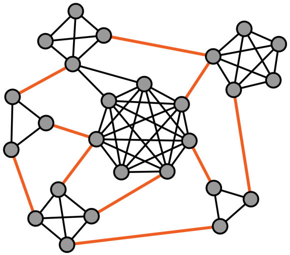
In this work, we compare the hierarchical and modular structure of clustered network models

Clustering

Maximally-ordered (CB) vs. Maximally-random (MR)

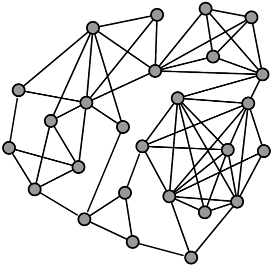
Maximally-ordered network (CB model)

- Fixed degree distribution, $p(k)$
- Fixed clustering distribution, $C(k)$
- Build a set of cliques
- Join cliques with configuration model



Maximally-random network (MR model)

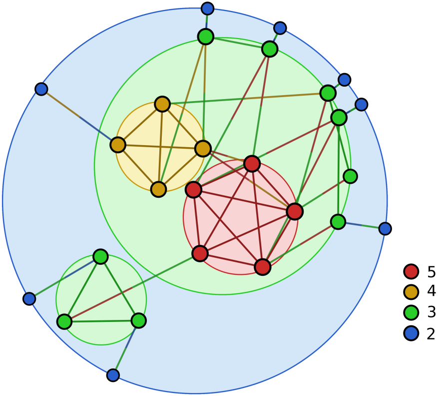
- Fixed degree distribution, $p(k)$
- Fixed clustering distribution, $C(k)$
- Define an exponential random graph
- Minimize Hamiltonian
  - Node rewiring
  - Simulated annealing



Clustering

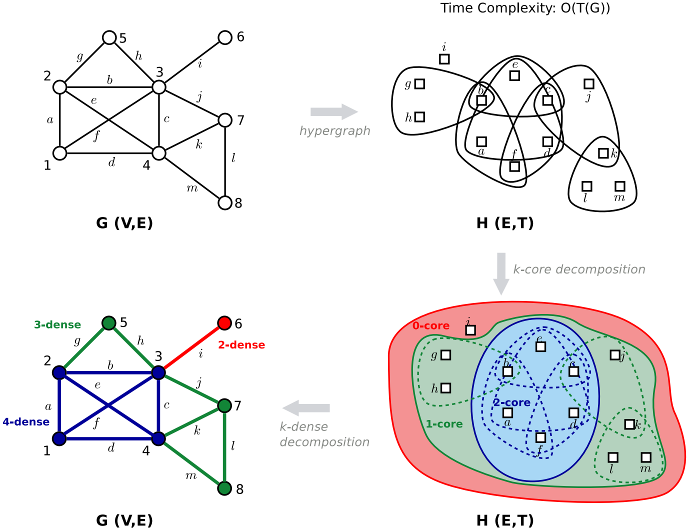
k-dense decomposition

- Based on the notion of edge multiplicity
- Edge multiplicity is the *number of common neighbors between the edge endpoints*
- A k-dense is a *maximal subgraph such that every edge has multiplicity at least* $(k-2)$ *inside the subgraph* [SYK08]



Clustering

k-dense decomposition: Algorithm



Clustering
Positioning algorithm

Combinatorial Models of Complex Systems

M.G. Beiró

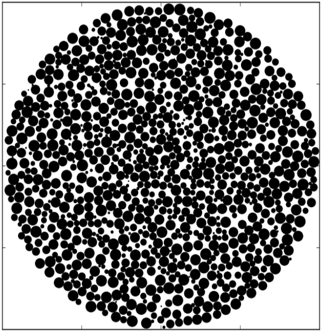
Introduction
Complex Systems
Example
Models of Complex Systems
Aims

Contributions
Internet
Clustering
Social Networks: Community Structure

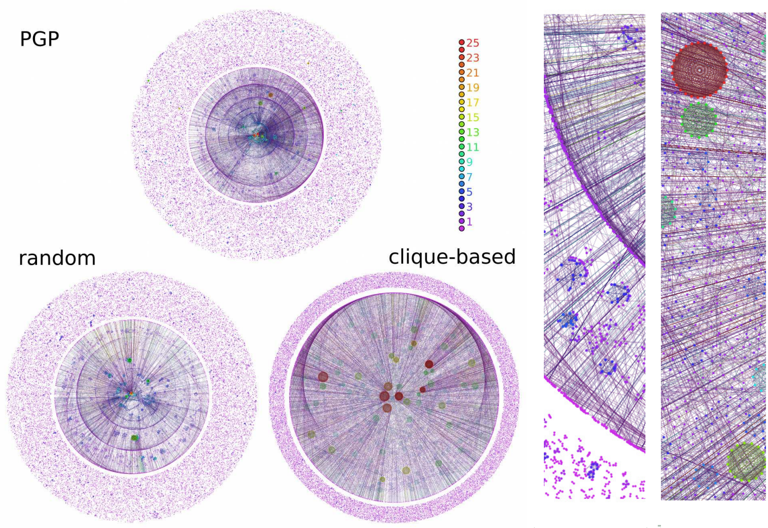
Extension: Community Discovery

Conclusions

- A connected component (CC) may split when we consider the $(k+1)$-dense
- Each CC is drawn into a circle
- Circles are randomly placed
- Random re-positioning until circles do not overlap

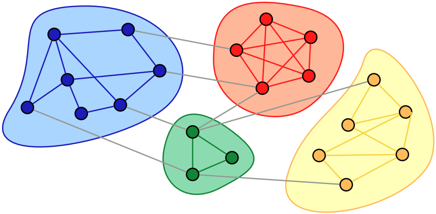


Clustering
PGP



Social Networks: Community Structure

- Community structure is the presence of groups of densily connected vertices
- Also called *modular structure*
- Found in many complex systems
  - Friendship networks, collaboration networks
  - The World Wide Web
  - Metabolic networks, networks of protein interactions
- But the concept lacks of a precise notion
  - Sometimes it depends on the application



- Notion of community
  - Concept of *web community* [FLG00]
  - Communities in the scientific collaboration network [New01]

- Some of the first methods:
  - Edge-betweenness, dendrogram [GN02]
  - Edge-clustering [RCC+04]
  - Modularity maximization
  - Label propagation (diffusion process) [RAK07]
  - INFOMAP (minimum description length) [RB08]
  - Lancichinetti's natural community (overlapping) [LFK09]

Social Networks: Community Structure

A local community detection method

Our method has 3 ingredients:

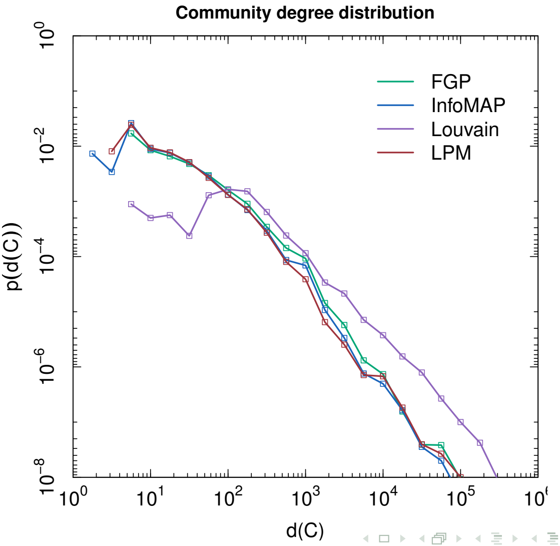
- A *fitness function* with a resolution parameter
- A *growth process*
  - The process "spreads" throughout the graph
  - The fitness function always increases
  - The resolution parameter is dynamically updated
- A *cutting technique*

The result:

- A *partition into communities*

The method was tested in:

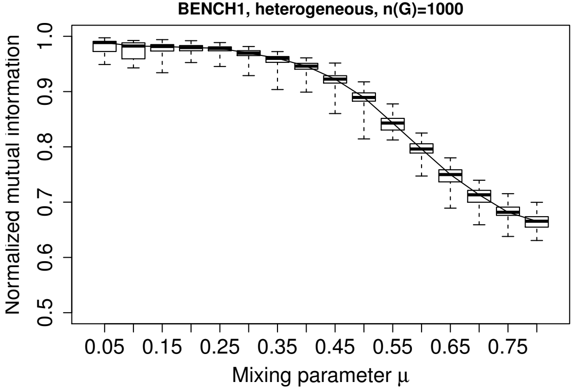
- Simple networks (karate, football, jazz, ...)
- Large real networks (web-Stanford (2M), LiveJournal (42M))
  - Power-law in the community size distribution
- LFR benchmark
  - Good performance in terms of the normalized mutual information


Community degree distribution
FGP
InfoMAP
Louvain
LPM
p(d(C))
d(C)
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^0
10^1
10^2
10^3
10^4
10^5

Social Networks: Community Structure

Results: Benchmark LFR

- Network instances with 1000 vertices
- Variable mixing parameters $[0{,}05 - 0{,}80]$



Topics

Our method has 3 ingredients:

- A *fitness function*
- A *growth process*
- A *cutting technique*

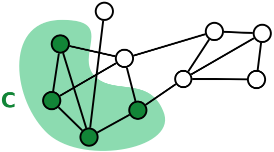
Its complexity is:

- Temporal: $O(n(G) \cdot d_{\max} + e(G) \cdot \log(n(G)))$
- Spatial: $O(n(G) + e(G))$

Given a subset $C \in V(G)$, we define the fitness function $H_t(C)$ as:

$$H_t(C) = m_V(C)\left(1 - \frac{m_V(C)}{2t}\right) - c_E(C) \qquad (1)$$

- What is $m_V(C)$?
  - Relative "mass" of C
- What is $c_E(C)$?
  - Relative external degree of C



- The growth process incorporates vertices into the set C, one after the other

- It always chooses the one which maximizes ΔH

- When no vertex can increase H we increase the resolution t as little as possible

What is the result?

- The growth process always prefers to incorporate vertices in the same local community
- When a community ends, we must incorporate a vertex from another community

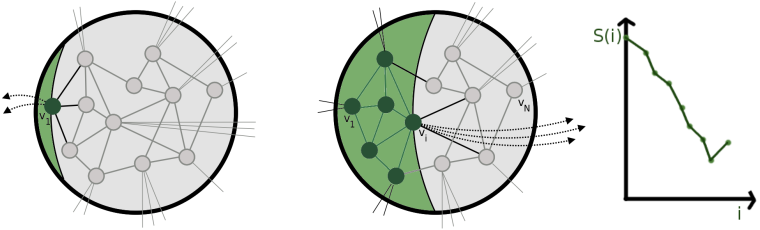
  Tipically, the resolution must be increased at this point
- Thus, the communities are traversed one after the other

Social Networks: Community Structure
Cutting technique

Each community evolves as $C_1, C_2, \ldots\ldots, C_{FINAL}$

We observe the behavior of the function:

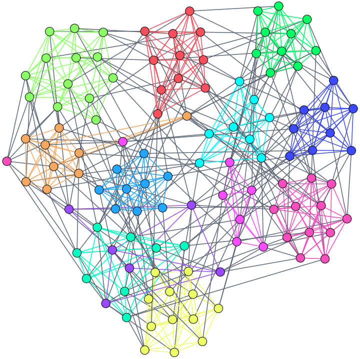
$$S(i) = \frac{c_E(C_i)}{m_V(C_i)} = \frac{\text{external connections}}{\text{community degree}}$$

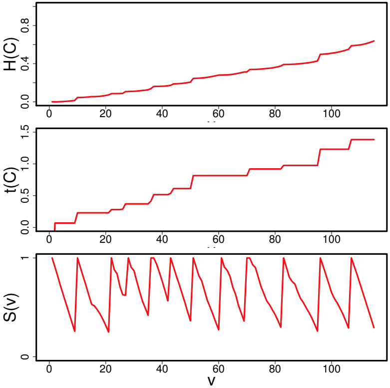


Social Networks: Community Structure

The growth process: An example

The “football” network:

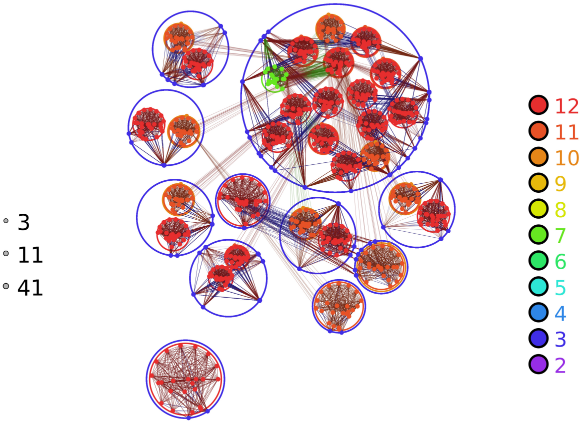




Social Networks: Community Structure

Community visualization

Stanford.edu web graph (250000 vertices) $\rightarrow$
Induced subgraph with **30 consecutive communities**
Node positioning according to the k-dense decomposition with LaNet-vi



(picture generated with SnailVis)

Topics

Conclusions

1. Connectivity in the Internet
   - Proposal estimator for edge-connectivity in the Internet based on the k-core decomposition
   - Edge-connectivity visualization with the LaNet-vi software
2. Clustering in networks
   - Efficient implementation of the k-dense decomposition
   - Development of a visualization tool for comparing clustered network models
3. Social networks: Discovery of community structure
   - Proposal of a local community detection algorithm
   - Visualization of community structure and evolution with the SnailVis software

Conclusions

- We performed a complexity analysis for each proposed algorithm
- We used visualization as a tool for understanding complex systems structure

Possible future lines of work:

- New criteria for evaluating quality of community structure
- Further applications of the k-dense decomposition

Thank you!

J.I. Alvarez-Hamelin, M.G. Beiró, and J.R. Busch.
Understanding edge connectivity in the internet through core decomposition.
Internet Mathematics, 7(1):45–66, 2011.

J.I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and Vespignani A.
k-core decomposition of internet graphs: hierarchies, self-similarity and measurement biases.
Networks and Heterogeneous Media, 3(2):371, 2008.

S. Carmi, S. Havlin, S. Kirkpatrick, and E. Shir.
A model of internet topology using k-shell decomposition.
PNAS, 104:11150–11154, 2007.

M. Faloutsos, P. Faloutsos, and C. Faloutsos.
On power-law relationships of the internet topology.
In *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, SIGCOMM '99, pages 251–262, New York, NY, USA, 1999. ACM.

G.W. Flake, S. Lawrence, and C.L. Giles.
Efficient identification of web communities.
In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '00, pages 150–160, New York, NY, USA, 2000. ACM.

M. Girvan and M.E.J. Newman.
Community structure in social and biological networks.
Proceedings of the National Academy of Sciences, 99(12):7821–7826, June 2002.

A. Lancichinetti, S. Fortunato, and J. Kertész.
Detecting the overlapping and hierarchical community structure in complex networks.
New Journal of Physics, 11(3):033015, 2009.

M.E.J. Newman.
The structure of scientific collaboration networks.
Proceedings of the National Academy of Sciences, 98(2):404–409, January 2001.

R. Pastor-Satorras, A. Vázquez, and A. Vespignani.
Dynamical and correlation properties of the internet.
Physical Review Letters, 87:258701+, 2001.

U.N. Raghavan, R. Albert, and S. Kumara.
Near linear time algorithm to detect community structures in large-scale networks.
Physical Review E, 76(3):036106+, September 2007.

M. Rosvall and C.T. Bergstrom.
Maps of random walks on complex networks reveal community structure.
Proceedings of the National Academy of Sciences, 105(4):1118–1123, 2008.

F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi.
Defining and identifying communities in networks.
Proceedings of the National Academy of Sciences, 101(9):2658, 2004.

G. Siganos, S.L. Tauro, and M. Faloutsos.
Jellyfish: A conceptual model for the as internet topology.
Journal of Communications and Networks, 8(3):339–350, 2006.

K. Saito, T. Yamada, and K. Kazama.
Extracting communities from complex networks by the k-dense method.
IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences, E91-A(11):3304–3311, November 2008.

D.J Watts and S.H. Strogatz.
Collective dynamics of small-world networks.
Nature, 393:440–442, 1998.