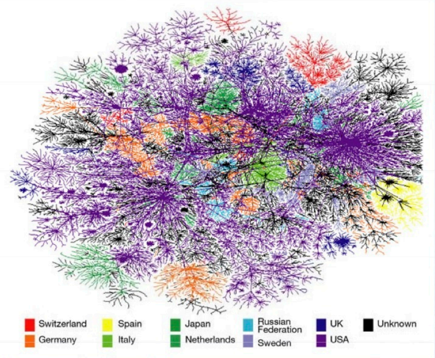
Reconstruction of Complex Networks

Francesc Comellas

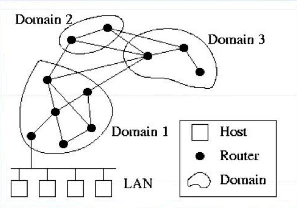
(joint work with J. Diaz-López and J.Paz)

Universitat Politècnica de Catalunya
Departament de Matemàtica Aplicada IV

1st Workshop on Graph-based Technologies and Applications (Graph-TA), Barcelona February 19, 2013

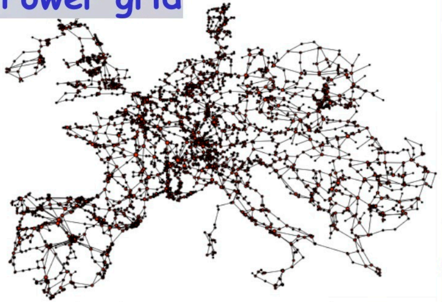


Internet

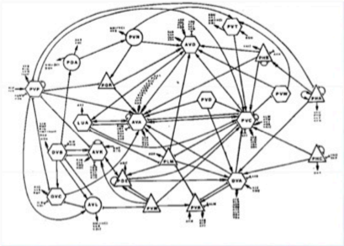


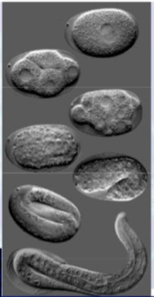
WWW

Power grid

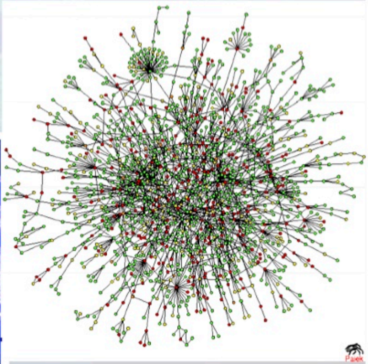


Air routes





C. Elegans



Proteins

Complex networks: transportation, communication (*WWW, Internet, P2P, AdHoc*), social, biological...

- Can be very large, even with millions of nodes, but usually they are sparse.
- Dynamical (P2P, AdHoc, ...) -their topology change with time.-

How would be possible to "store" them compactly ?

Can we create similar networks given some (a limited number) parameters?

Use graph invariants !

Order, size, radius, mean distance, average degree, chromatic number, covering number, degree sequence, eccentricity sequence, betweenness centrality, **spectrum (adjacency, laplacian,...)**, etc.

For example, if we know that a certain graph has this **betweenness centrality** (a measure of the load of its vertices: fraction of all shortest paths that go through a given node):

$$\{0, \frac{2}{5}, \frac{3}{5}, , 0, 0, 0\}$$

or if we know that a graph has as its **Laplacian spectrum**: (eigenvalues of the Laplacian matrix of the graph $L = A - D$):

$$\{0, 1, 1, 3, 3, 6\}$$

Can we find the graphs ?
Is there only one graph with this BC / spectrum or are there many or none ?

From the spectrum to the graph

Spectrum: $\{0, 1, 1, 3, 3, 6\}$

(Eigenvalues of the Laplacian matrix, $L = A - D$, of the graph.)

We know that the graph has

- Order $n = 6$.
- $\sum_i \lambda_i = 14 = \sum_v deg(v) = 2|E|$. 7 edges.
- $\Delta \leq \frac{n-1}{n}\lambda_n = 5$. $\delta \geq \frac{n-1}{n}\lambda_2 = 0.8333$.
- Not regular.

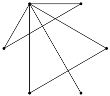
Can we allways find the graph ?

From the spectrum to the graph

Spectrum: $\{0, 1, 1, 3, 3, 6\}$

(Eigenvalues of the Laplacian matrix, $L = A - D$, of the graph.)

We know that the graph has

- Order $n = 6$.
- $\sum_i \lambda_i = 14 = \sum_v deg(v) = 2|E|$. 7 edges.
- $\Delta \leq \frac{n-1}{n}\lambda_n = 5$. $\delta \geq \frac{n-1}{n}\lambda_2 = 0.8333$.
- Not regular.
- 

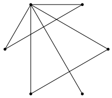
Can we allways find the graph ?

From the spectrum to the graph

Spectrum: $\{0, 1, 1, 3, 3, 6\}$

(Eigenvalues of the Laplacian matrix, $L = A - D$, of the graph.)

We know that the graph has

- Order $n = 6$.
- $\sum_i \lambda_i = 14 = \sum_v deg(v) = 2|E|$. 7 edges.
- $\Delta \leq \frac{n-1}{n}\lambda_n = 5$. $\delta \geq \frac{n-1}{n}\lambda_2 = 0.8333$.
- Not regular.
-  Problem: There are 156 graphs with 6 vertices, and 24 graphs among them with these same properties, but only one with this exact spectrum. However, for $n = 12$ there are millions of graphs for a given size and some bounds in the degrees.

Can we allways find the graph ?

From the betweenness centrality to a graph

The betweenness centrality of vertex w is $\beta_w = \frac{1}{(n-1)(n-2)} \sum_{u,v \neq w} \frac{\sigma_{uv}(w)}{\sigma_{uv}}$

BC $\{0, \frac{2}{5}, \frac{3}{5}, , 0, 0, 0\}$

- The graph has 6 vertices. There are 156 graphs with 6 vertices.

From the betweenness centrality to a graph

The betweenness centrality of vertex w is $\beta_w = \frac{1}{(n-1)(n-2)} \sum_{u,v \neq w} \frac{\sigma_{uv}(w)}{\sigma_{uv}}$

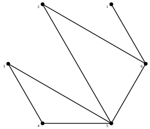
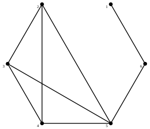
BC $\{0, \frac{2}{5}, \frac{3}{5}, , 0, 0, 0\}$

- The graph has 6 vertices. There are 156 graphs with 6 vertices.
- Edges ? Degrees ? Connected ? Is it regular ? ..

From the betweenness centrality to a graph

The betweenness centrality of vertex w is $\beta_w = \frac{1}{(n-1)(n-2)} \sum_{u,v \neq w} \frac{\sigma_{uv}(w)}{\sigma_{uv}}$

BC $\{0, \frac{2}{5}, \frac{3}{5}, , 0, 0, 0\}$

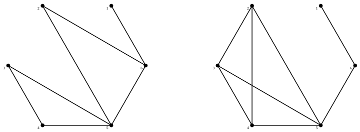
- The graph has 6 vertices. There are 156 graphs with 6 vertices.
- Edges ? Degrees ? Connected ? Is it regular ? ..
-  

  Problem There are two graphs with this BC and they are not isomorphic.

From the betweenness centrality to a graph

The betweenness centrality of vertex w is $\beta_w = \frac{1}{(n-1)(n-2)} \sum_{u,v \neq w} \frac{\sigma_{uv}(w)}{\sigma_{uv}}$

BC $\{0, \frac{2}{5}, \frac{3}{5}, , 0, 0, 0\}$

- The graph has 6 vertices. There are 156 graphs with 6 vertices.
- Edges ? Degrees ? Connected ? Is it regular ? ..
- 

  Problem There are two graphs with this BC and they are not isomorphic.
- In fact all (156) graphs with 6 vertices have diferent BC, except this an another pair !

Cospectral graphs.

cospectral with respect to the Laplacian matrix $L = D - A$

Two graphs with the same spectrum but topologically distinct.

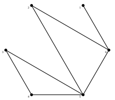
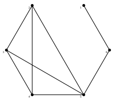
$|V| < 6$ There are no cospectral graphs.
$|V| = 6$ there are 4 cospectral graphs (out from 156 graphs).
$|V| = 7$ there are 130 cospectral graphs (out from 1044 graphs).
$|V| = 8$ there are 1767 cospectral graphs (out from 12346 graphs), etc.

The ratio cospectral vs. total number of graphs goes to zero with the order.

E.R. van Dam, W.H. Haemers. E. Spencer

There are too many possible graphs to explore (even with the rectrictions that come from theoreical studies). NP-complete problem ! Combinatorial optimization methods are useful.

Graphs with the same BC.

Non isomorphic with the same betweenness centrality.

$|V| < 6$ there are no graphs.
$|V| = 6$ there are 2 pairs (out from 156 graphs).
$|V| = 7$ there are 15 pairs (out from 1044 graphs).
$|V| = 8$ there are 92 pairs (out from 12346 graphs).

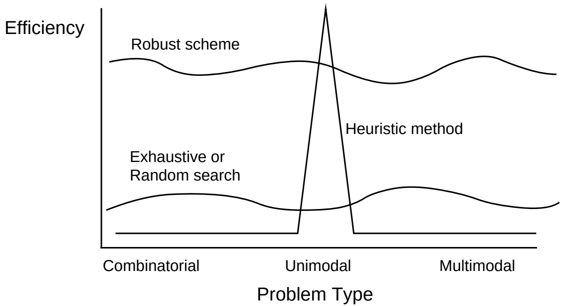
The number of on isomorphic graphs with the same BC seems to increase rapidly with the order of the graph, but the fraction is very small.

Hence, two graphs with the same BC would indeed be isomorphic with a high probability.

Given a BC which corresponds to a graph, is it an NP-complete problem to find it?

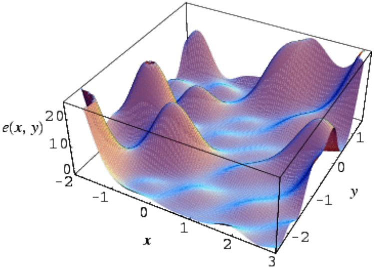
Combinatorial optimization methods will certainly be useful !

Efficiency *versus* effectiveness



We look for robust schemes such that they may be used for a wide range of problems:

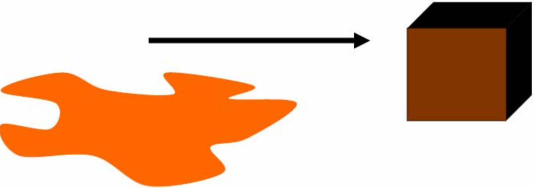
The main problem is to find the global maximum.
We should prove that, with a high probability, the method finds solutions not far from this maximum.



E. Aarts, K. Lenstra, (Eds.), *Local Search in Combinatorial Optimization.* John Wiley & Sons. Ltd. Chichester, 1997.

Simulated annealing / Metropolis algorithm

Classical meethod inspired by cristalization processes.



- N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller, *J. Chem. Phys.*, **21**, 1087, 1953.
- S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, Optimization by Simulated Annealing. *Science*, **220**, 671-680, 1983.

Simulated annealing

1. Generate an initial random graph. Fix T_0 and T_{min}.
2. Repeat N_k times.
    1. Modify the graph and find new cost.
    2. If better, accept it as current solution.
    3. If worse, accept only if $e^{-\Delta\epsilon/KT_k} > rand()$
3. Lower T_k and repeat 2 until $T_k < T_{min}$

Important parameters to adjust are: T_0, N_k, and the cooling rate α from $T_{k+1} = \alpha T_k$.

Tabu search

The main idea is to perform a greedy search avoiding repetition of steps. Keeps a list which is updated at every step removing the oldest elements and allowing a cost improvement and thus local minima.

- F. Glover. Future paths for integer programming and links to artificial intelligence. *Comput. & Ops. Res.* **13** (1986) 533–549.

Tabu search

1. Generate an initial random graph. Tabu list empty.
2. Repeat until stop criterion
    1. Select vertex at random and modify an edge not in tabu list. Compute cost.
    2. If better, accept graph and add edge to tabu list.
    3. If worse, undo modification and add edge to tabu list
    4. Delete old items from tabu list.

The main parameter to adjust is the tabu list size. Note that because of this list, modifications that improve the cost function can be rejected.

Algorithm design.

- Generate an initial random graph with:
  - Order n
  - Size m

    random $n-1 \leq m \leq n(n-1)$ for BC reconstruction.

    $m = \sum_i \lambda_i/2$ for spectral reconstruction.
  - Connected.

    Maximum and minimum degrees can be obtained from the spectrum in the case of spectral reconstruction.
- Modification performed at each iteration step:

  Link reconnections.
- Verify the “quality” of the graph (cost function)
- Compare the final graph with the reference (initial) graph.

Cost function for spectral reconstruction

The best cost function is:

$$\sqrt{\sum_{i=0}^{n}(\lambda_i^0 - \lambda_i^t)^2}$$

Where λ_i^0, $1 \leq i \leq n$ are the Laplacian eigenvalues of the reference graph and λ_i^t those of the graph being tested.

We have also tested the following cost functions:

- $\sqrt{\sum_{i=1}^{n}(i+1)(\lambda_i^0 - \lambda_i^t)^2}$,
- $\sqrt{\sum_{i=1}^{n}(n-i)(\lambda_i^0 - \lambda_i^t)^2}$,
- $\sqrt{\sum_{i=1}^{n}|\lambda_i^0 - \lambda_i^t|^3}$,
- $\sqrt{\sum_{i=1}^{n}|\lambda_i^0 - \lambda_i^t|(n-1)^{1.5}}$.

Cost function for BC

We use

$$\epsilon = \sum_{i=1}^{n} (\beta_i^0 - \beta_i^t)^2.$$

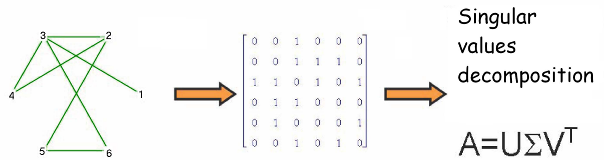
where $\{\beta_1^0, \beta_2^0, \ldots, \beta_n^0\}$ is the BC of the reference graph and $\{\beta_1^t, \beta_2^t, \ldots, \beta_n^t\}$ is that of the graph being tested.

Other cost functions considered:

- $\sqrt{\sum_{i=0}^{n} (\beta_i^0 - \beta_i^t)^2}$,
- $\sqrt{\sum_{i=0}^{n} (i+1)(\beta_i^0 - \beta_i^t)^2}$,
- $\sqrt{\sum_{i=0}^{n} (n-i)(\beta_i^0 - \beta_i^t)^2}$,
- $\sqrt{\sum_{i=0}^{n} |\beta_i^0 - \beta_i^t|^3}$,
- $\sqrt{\sum_{i=0}^{n} |\beta_i^0 - \lambda_i^t|(n-1)^{1.5}}$.

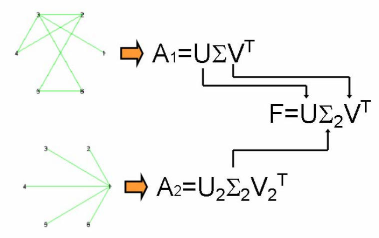
How we can measure a distance between graphs ?
Singular values descomposition method.

SVD method introduced in M. Ipsen, A.S. Mikhailov..



A is the adjacency matrix of the graph.
U and V are unitary matrices.
Σ is the diagonal matrix of singular values.

Calculation of a *distance* between graphs.



Graphs G_1, G_2 with adjacency matrices A_1, A_2.
$F = F(A_1, A_2) = U_1 \Sigma_2 V_1^T = U_1 U_2^T A_2 V_2 V_1^T$.

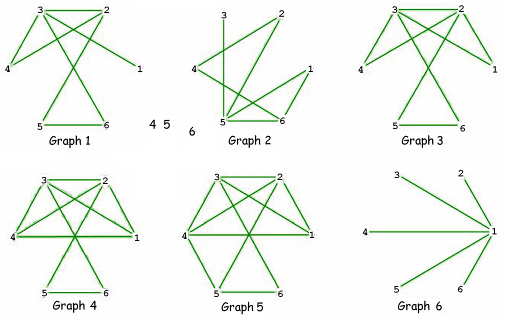
If the graphs are isomorphic then $A_1 = F(A_1, A_2)$.

For a "good" reconstruction, F will have real values not far from the 1s and 0s of A_1.
Introducing $\Delta = A_1 - F$ and the norm

$$\delta = \frac{1}{n} \sqrt{\sum_{i,j} \Delta_{ij}^2}$$

we obtain a measure for the distance between the two graphs.

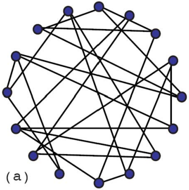
An example with values for δ



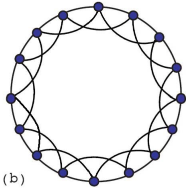
	G2	G3	G4	G5	G6
δ	0	0,134	0,160	0,167	0,474

- C++ compiler DEV-C++
- PC (P4 CPU at 2.41 GHZ) Windows XP.
- Limited to 300 s per reconstruction.
- Average of 100 (Sp) - 500 (BC) reconstructions.

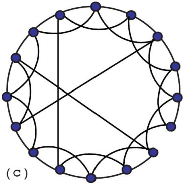
Initial Graphs



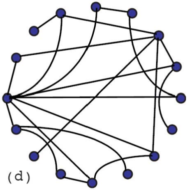
Random



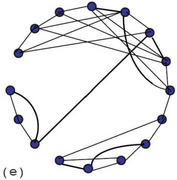
Circulant



Small-World (WS)

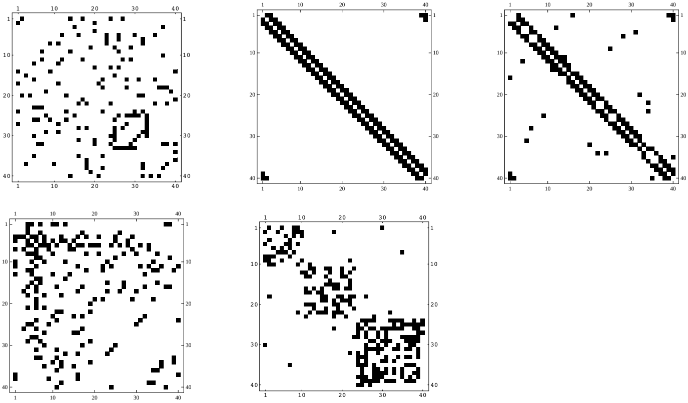


Scale-Free (BA)



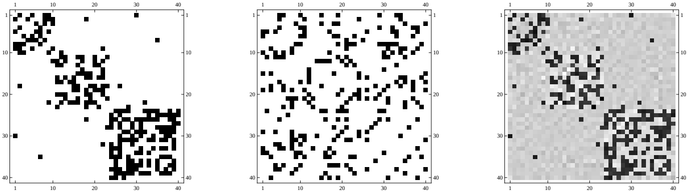
Clustered

Adjacency matrices of the reference graphs

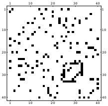
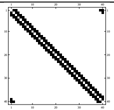
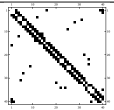
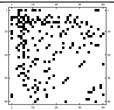
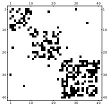


A visualization for a reference graph and its reconstruction F

An example: Spectral reconstruction of a clustered graph using tabu search.



Adj. matrix (original) - Adj. matrix (reconstr.) - Matrix F (reconstr.)

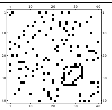
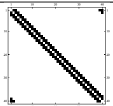
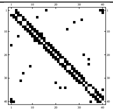
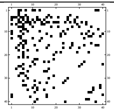
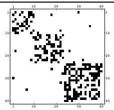
		 Random		 Circulant		 Small-world		 Scale-free		 Clustered	
		Ref.	*Recn.*	Ref.	*Recn.*	Ref.	*Recn.*	Ref.	*Recn.*	Ref.	*Recn.*
Diameter	avg.	6	6.4	12	9.7	9	7.9	4	4.1	5	5
Avg. dist.	avg.	2.89	2.9	6.19	4.22	3.97	3.72	2.32	2.32	2.67	2.23
Degrees	min.	1	1	4	1	2	1	1	1	2	1
	avg.	3.8	3.8	4	4	3.9	3.9	4.95	4.95	6.3	6.3
	max.	8	8	4	5	6	6	17	17	13	13
Clustering	avg.	0.20	0.22	0.7	0.73	0.53	0.49	0.23	0.19	0.37	0.22
δ	avg.		0.02		0.08		0.04		0.02		0.08

Tabu search.

Results for the average of 100 reconstructions for each reference graph.

Spectrum (graph order): 40. Eigenvalue tolerance: 0.0001. Tabu list size: 400.

Tabu iterations after a change: 20.

		 Random		 Circulant		 Small-world		 Scale-free		 Clustered	
		Ref.	*Recn.*	Ref.	*Recn.*	Ref.	*Recn.*	Ref.	*Recn.*	Ref.	*Recn.*
Diameter	avg.	6	5.87	10	8.77	6	6.69	4	4.48	5	5.27
Avg. Dist.	avg.	2.89	2.78	5.38	4.67	3.31	3.03	2.32	2.34	2.65	2.58
Degrees	min.	1	1.00	4	1.99	3	1.00	1	1.11	2	3.75
	avg.	3.8	3.95	4	2.57	4	3.69	4.95	5.10	6.3	4.46
	max.	8	10.00	4	3.87	5	6.19	17	15.76	13	14.65
Clustering	avg.	0.2	0.20	0.5	0.03	0.32	0.12	0.26	0.26	0.37	0.20
Norm. BC	avg.	0.05	0.05	0.11	0.10	0.06	0.05	0.04	0.04	0.04	0.04
δ	avg.		0.03		0.07		0.03		0.02		0.08

Simulated annealing.

Results for the average of 500 reconstructions for each reference graph. Graph order$= 40$, $T_0 = 1.0$, $N = 2000$, $T_{min} = 0.000001$, geometric cooling rate $T_{k+1} = 0.9T_k$.

Conclusion

Given the Laplacian spectrum, or the betweenness centrality (BC), the methods allows an exact reconstruction for graphs with less than 14 vertices.

It is possible a “topological” reconstruction for any graph (tests with graphs of order 2000 and size 20000).

Even if the spectrum does not correspond to a graph, the method constructs a similar network topology. This can be used to generate network models or even real topologies with specific properties as traffic / load distributions (hubs, terminal nodes, etc.)