Updating and downdating techniques for optimizing network communicability

Francesca Arrigo and Michele Benzi

University of Insubria (Como, Italy)
Emory University (Atlanta, USA)

CIME–EMS Summer School, Cetraro
June 23, 2015

Complex networks. . . ?

A complex network is a graph that exhibits non-trivial topological features.

A complex network is not a regular, nor a random graph.

They are used to model interactions of various types:

- social networks: collaboration, friendship,. . . ;
- biological networks: PPI, food webs,. . . ;
- technological networks: www, internet,. . . ;
- transportation network: air routes, road maps,. . .

and of course they are used to describe

- Facebook, Twitter, Instagram, . . .

Complex networks. . . ?

A complex network is a graph that exhibits non-trivial topological features.

A complex network is not a regular, nor a random graph.

They are used to model interactions of various types:

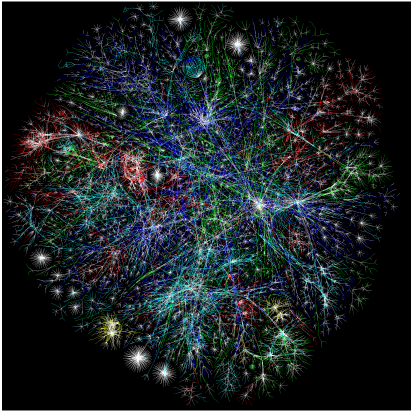
- social networks: collaboration, friendship,. . . ;
- biological networks: PPI, food webs,. . . ;
- technological networks: www, internet,. . . ;
- transportation network: air routes, road maps,. . .

and of course they are used to describe

- Facebook, Twitter, Instagram, . . .

Motivation

We aim at making (minimal) modifications to an existing network so as to maximize its ability to propagate "information" along its edges.



Total communicability (TC):

quantifies the ease of spreading information across the network and how well connected a network is.

We want to manipulate the edges in the network in order to tune the TC.

Background

Let $G = (W, E)$ be a complex network with $n = |W|$ nodes and $m = |E|$ edges.
Suppose that:

- G is unweighted;
- G has no multiple edges nor self loops;
- $m = O(n)$.

Adjacency matrix

G can be represented using an *adjacency matrix* $A \in \mathbb{R}^{n \times n}$

$$(A)_{ij} = a_{ij} = \begin{cases} \omega_{ij} & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

where $\omega_{ij} \in \mathbb{R}$ are weights for the elements in E.

In our framework $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ will be:

- binary and $a_{ii} = 0$ for all $i \in V$;
- symmetric, if G is undirected;
- sparse.

A few useful definitions when $A^T = A$

- **walk** of length k:
  $\{i_1, i_2, \ldots, i_{k+1} \in V | (i_l, i_{l+1}) \in E \text{ for all } 1 \leq l \leq k\}$
- **closed walk** of length k: a walk for which $i_1 = i_{k+1}$.

Remark

The quantities $\left(A^k\right)_{ii}$, $\left(A^k\right)_{ij}$ count closed (resp., open) walks of length k.

- **degree** of node i:

$$d_i = |\{j \in V : (i, j) \in E\}| = \left(A^2\right)_{ii}.$$

- **total communicability** of an undirected network

$$TC(A) = \mathbf{1}^T e^A \mathbf{1} = \sum_{k=0}^{\infty} \mathbf{1}^T \frac{A^k}{k!} \mathbf{1}$$

A few useful definitions when $A^T = A$

- **walk** of length k:
  $\{i_1, i_2, \ldots, i_{k+1} \in V | (i_l, i_{l+1}) \in E \text{ for all } 1 \leq l \leq k\}$
- **closed walk** of length k: a walk for which $i_1 = i_{k+1}$.

Remark

The quantities $\left(A^k\right)_{ii}$, $\left(A^k\right)_{ij}$ count closed (resp., open) walks of length k.

- **degree** of node i:

$$d_i = |\{j \in V : (i,j) \in E\}| = \left(A^2\right)_{ii}.$$

- **total communicability** of an undirected network

$$\boxed{TC(A) = \mathbf{1}^T e^A \mathbf{1} = \sum_{k=0}^{\infty} \mathbf{1}^T \frac{A^k}{k!} \mathbf{1}}$$

Problem setting

We develop techniques aimed at tackling the following problems.

- Downdate: select K edges that can be downdated from the network that cause the smallest drop in $TC(A)$;
- Update: select K virtual edges to be added to the network so as to increase as much as possible the total communicability of the graph;

Edges connecting important nodes are themselves important:

$$^{e}C(i,j) = C(i) \cdot C(j), \qquad \forall i, j \in V$$

where $C(\cdot) : V \to \mathbb{R}$ is a centrality measure for nodes.

Node centrality measures

The centrality measures we use are all walk–based:

1. *eigenvector centrality*:

$$EC(i) = \mathbf{q}_1(i);$$

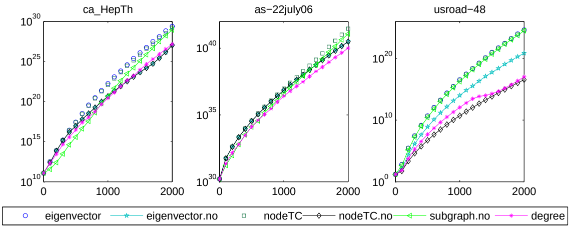
2. *subgraph centrality*:

$$SC(i) = (e^A)_{ii} = \sum_{k=0}^{\infty} \frac{(A^k)_{ii}}{k!};$$

3. *total communicability*:

$$TC(i) = (e^A\mathbf{1})_i = \sum_{k=0}^{\infty} \frac{(A^k\mathbf{1})_i}{k!}.$$

UP: Results & Timings (in sec.)



	ca–HepTh	as–22july06	usroad–48
eigenvector	192.8	436.9	1599.5
nodeTC	561.94	1218.77	2932.01
degree	11.1	12.4	175.8
eigenvector.no	**0.19**	**0.33**	**5.85**
nodeTC.no	**0.30**	**0.55**	**1.59**
subgraph.no	10.6	74.1	651.6

The directed case: $A^T \neq A$

Nodes play two different roles in digraphs:

broadcasters of information (hubs)

receivers of information (authorities)

We need to define two quantities that describe the overall ability of the nodes in the digraph to play these roles and that reflect the following ideas:

good hubs are nodes that point to many good authorities.

good authorities are nodes that are pointed to by many good hubs.

Alternating walks

As for the symmetric case, we want to "count walks", respecting the recursive definition that relates hubs and authorities.

Definition

Alternating walk of length k starting with and out-edge

$$\{i_1, i_2, \ldots, i_{k+1} | (i_\ell, i_{\ell+1}) \in E \textit{ iff } \ell \equiv_2 1;\ (i_{\ell+1}, i_\ell) \in E \textit{ iff } \ell \equiv_2 0\}.$$

$$i_1 \rightarrow i_2 \leftarrow i_3 \rightarrow \ldots.$$

- $\left[(AA^T)^k\right]_{ij}$ counts alternating walks of length $2k$ starting with an out-edge between i and j.
- AA^T in called HUB MATRIX.

Alternating walks (cont.)

As for the symmetric case, we want to "count walks", respecting the recursive definition that relates hubs and authorities.

Definition

Alternating walk of length k starting with and in-edge

$$\{i_1, i_2, \ldots, i_{k+1} | (i_\ell, i_{\ell+1}) \in E \textit{ iff } \ell \equiv_2 0;\ (i_{\ell+1}, i_\ell) \in E \textit{ iff } \ell \equiv_2 1\}.$$

$$i_1 \leftarrow i_2 \rightarrow i_3 \leftarrow \ldots.$$

- $\left[(A^T A)^k\right]_{ij}$ counts alternating walks of length $2k$ starting with an in-edge between i and j.
- $A^T A$ in called AUTHORITY MATRIX.

Total hub/authority communicability

To quantify the ability of the network to broadcast information when all the nodes are acting as hubs we use the total hub communicability:

$$T_h C(A) := \mathbf{1}^T \left(\sum_{k=0}^{\infty} \frac{(AA^T)^k}{(2k)!} \right) \mathbf{1} = \mathbf{1}^T \cosh(\sqrt{AA^T}) \mathbf{1}.$$

To quantify the ability of the network to receive information when all the nodes are acting as authorities we use the total authority communicability:

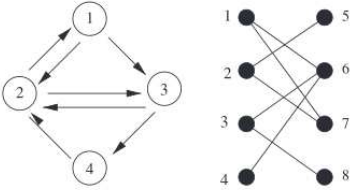
$$T_a C(A) := \mathbf{1}^T \left(\sum_{k=0}^{\infty} \frac{(A^T A)^k}{(2k)!} \right) \mathbf{1} = \mathbf{1}^T \cosh(\sqrt{A^T A}) \mathbf{1}.$$

Digraphs as bipartite networks

Let $\mathcal{G} = (\mathcal{W}, \mathcal{E})$ be a graph such that

$$\begin{aligned} \mathcal{W} &= W \cup W' \\ W' &= \{1' = n+1, 2' = n+2, \ldots, n' = 2n\} \\ &= \text{set of copies of the nodes} \\ \mathcal{E} &= \{(i, j') : (i, j) \in E\}. \end{aligned}$$

The associate adjacency matrix is $\mathcal{A} = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix}$.



Digraphs as bipartite networks (cont.)

The matrix exponential of the adjacency matrix $\mathcal{A}$ is:

$$e^{\mathcal{A}} = \begin{pmatrix} \cosh(\sqrt{AA^T}) & \mathrm{gsinh}(A) \\ [\mathrm{gsinh}(A)]^T & \cosh(\sqrt{A^T A}) \end{pmatrix},$$

where

$$\mathrm{gsinh}(A) := U \sinh(\Sigma) V^T,$$

and $A = U\Sigma V^T$ is the SVD of the adjacency matrix.

Heuristics

To tune the indices of interest we can select the modification (i, j) to be performed by working:

(1) on the symmetric matrix $\mathcal{A}$ using the techniques previously introduced;

(2) on the original matrix A.

Edges connecting important nodes are themselves important:

$$^{e}C(i,j) = C_h(i) \cdot C_a(j), \qquad \forall i, j \in V$$

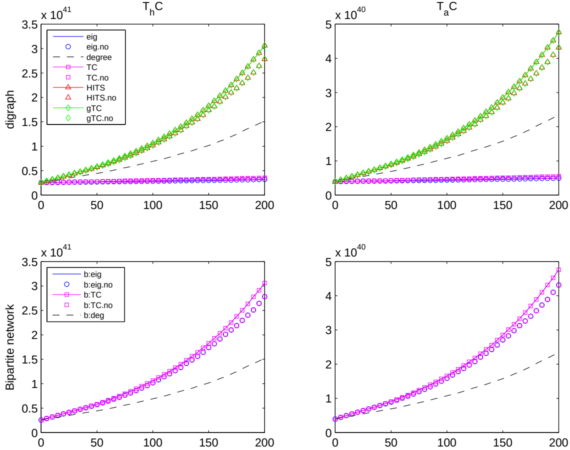
where $C_h(i)$ and $C_a(i)$ are the centrality as hub and authority of node i, respectively.

Remark

The set where we search for the modifications is the same, regardless of whether we are working on $\mathcal{A}$ or A.

An example: UPDATE ($K = 200$)

Network:	n	m	σ_1	σ_2	$\sigma_1 - \sigma_2$
cit-HepTh	27770	352807	85.16	69.31	15.85



Conclusions

- We have developed efficient and effective heuristics for optimizing the network communicability, for both directed and undirected networks;
- the leading eigenvalue λ_1 (resp., singular value σ_1) and the spectral gap $\lambda_1 - \lambda_2$ (resp., $\sigma_1 - \sigma_2$) play a major role in the evolution of the indices and give insights on the "agreement" of the methods;
- in the undirected case, our methods scale linearly in practice;
- it is fundamental to use edge centrality measures;
- in the undirected case, the TC can be used as a measure of network connectivity.

We still need to assess the scalability of our methods in the directed case!

- -Thank you!

Conclusions

- We have developed efficient and effective heuristics for optimizing the network communicability, for both directed and undirected networks;
- the leading eigenvalue λ_1 (resp., singular value σ_1) and the spectral gap $\lambda_1 - \lambda_2$ (resp., $\sigma_1 - \sigma_2$) play a major role in the evolution of the indices and give insights on the "agreement" of the methods;
- in the undirected case, our methods scale linearly in practice;
- it is fundamental to use edge centrality measures;
- in the undirected case, the TC can be used as a measure of network connectivity.

We still need to assess the scalability of our methods in the directed case!

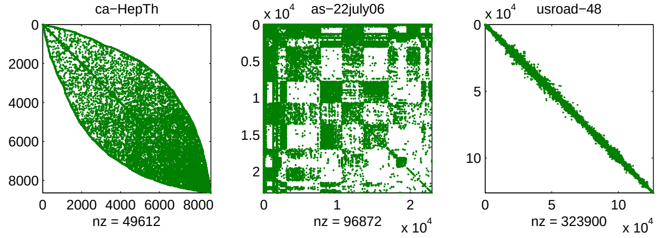
- -Thank you!

Bibliography

- F. Arrigo & M. Benzi, *"Updating and downdating techniques for optimizing network robustness"*, submitted;
- M. Benzi, E. Estrada, and C. Klymko, *Ranking hubs and authorities using matrix functions*, Linear Algebra Appl. 438 (2013), pp. 2447–2474
- J. B. Hawkins & A. Ben–Israel, *On generalized matrix functions*, Linear and Multilinear Algebra 1(2) (1973), pp. 163–171.
- N. J. Higham, *Function of Matrices. Theory and Computation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2008.
- J. Kleinberg, *Authoritative sources in a hyperlinked environment*, J. ACM 46 (1999), pp. 604–632.

Dataset

NAME	n	m	λ_1	λ_2	$\lambda_1 - \lambda_2$
ca–HepTh	8638	24806	31.034	23.004	8.031
as–22july06	22963	48436	71.613	53.166	18.447
usroad–48	126146	161950	3.911	3.840	0.071



Dataset

NAME	n	m	λ_1	λ_2	$\lambda_1 - \lambda_2$
ca–HepTh	8638	24806	31.034	23.004	8.031
as–22july06	22963	48436	71.613	53.166	18.447
usroad–48	126146	161950	3.911	3.840	**0.071**