

# Ranking nodes in signed networks: an algebraic perspective

**Dmitry Gromov**

University of Ostrava  
April 24, 2023



UNIVERSITY  
OF LATVIA

# Complex networks

*Complex networks* are ubiquitous:

- Social networks
- Biological networks
- Computer networks
- Communication networks
- Electrocatal networks
- et cetera ...



©Martin Grandjean



UNIVERSITY  
OF LATVIA

# Complex networks

Complex networks are closely related to *graphs*.

However, while graphs are *deterministic* objects, complex networks are mostly studied using *statistical* methods.



©Martin Grandjean



UNIVERSITY  
OF LATVIA

# Graphs

A *weighted graph* is described by a tuple

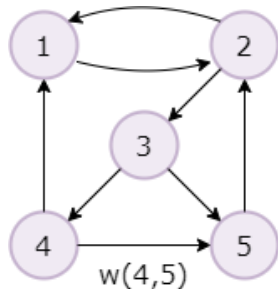
$$\Gamma = (V, E, w),$$

where  $V = \{1, \dots, N\}$  is the set of vertices,

$E \subset V \times V$  the set of edges,

$w : E \rightarrow \mathbb{R} \setminus \{0\}$  the weighting function.

A weighted graph is said to be *unsigned* if the weighting function maps to  $\mathbb{R}_+$  and *signed* if  $w$  maps to  $\mathbb{R} \setminus \{0\}$ .



An important characteristic of a network is the *degree distribution*.

$$d^{in}(\Gamma) = \{2, 2, 1, 1, 2\}$$

$$d^{out}(\Gamma) = \{1, 2, 2, 2, 1\}$$



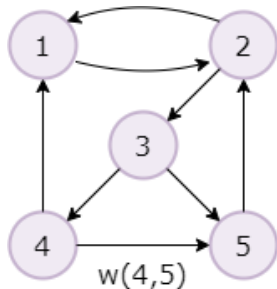
# Graphs

A *weighted graph* is described by a tuple

$$\Gamma = (V, E, w),$$

where  $V = \{1, \dots, N\}$  is the set of vertices,  
 $E \subset V \times V$  the set of edges,  
 $w : E \rightarrow \mathbb{R} \setminus \{0\}$  the weighting function.

A weighted graph is said to be *unsigned* if the weighting function maps to  $\mathbb{R}_+$  and *signed* if  $w$  maps to  $\mathbb{R} \setminus \{0\}$ .



An important characteristic of a network is the *degree distribution*.

$$\mathbf{d}^{in}(\Gamma) = \{2, 2, 1, 1, 2\}$$

$$\mathbf{d}^{out}(\Gamma) = \{1, 2, 2, 2, 1\}$$

# Degree distribution in complex networks

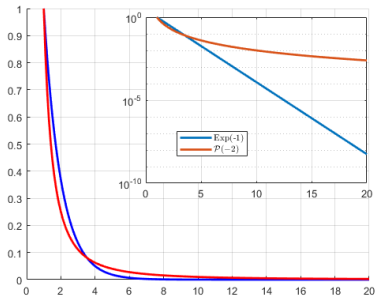
The degree distribution of most **real** complex networks is described by a *power law*:

$$X \sim \mathcal{P}(k) : f(x) = ax^{-k}, \quad k > 1.$$

In most cases,  $2 < k < 3$ .

Some properties:

- The mean is defined for  $k > 2$ .
- Finite variance for  $k > 3$ .
- Heavy tail, finite probability of extreme events.



# Degree distribution in complex networks

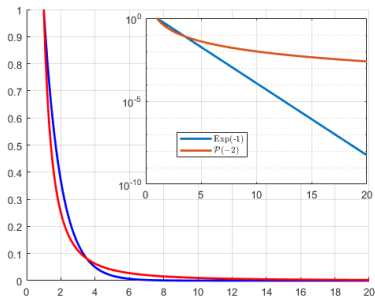
The degree distribution of most **real** complex networks is described by a *power law*:

$$X \sim \mathcal{P}(k) : f(x) = ax^{-k}, \quad k > 1.$$

In most cases,  $2 < k < 3$ .

Some properties:

- The mean is defined for  $k > 2$ .
- Finite variance for  $k > 3$ .
- Heavy tail, finite probability of extreme events.



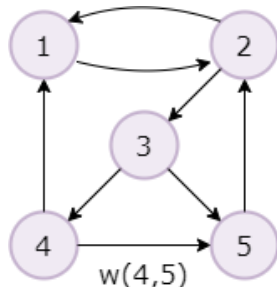
# Algebraic description

A weighted graph  $\Gamma$  is uniquely described by its adjacency matrix  $A$ .

$$a_{ij} = \begin{cases} w(j, i), & (i, j) \in E, \\ 0, & \text{otherwise.} \end{cases}$$

For the graph shown in Fig.:

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$





# Elementary algebraic operations

Define  $\mathbf{1} = [1, 1, \dots, 1]^T$  to be the column vector of all ones. Then,

$$\mathbf{d}^{out} = \mathbf{1}^T \cdot \mathbf{A}$$

$$\mathbf{d}^{in} = \mathbf{A} \cdot \mathbf{1}.$$

Furthermore, we define the square matrices  $\mathbf{D}^{out} = \text{diag}(\mathbf{d}^{out})$  and  $\mathbf{D}^{in} = \text{diag}(\mathbf{d}^{in})$ .



















# First remedy: normalization

To overcome this difficulty, we assume that the score is being split equally between all outbound nodes.

Thus, we modify the adjacency matrix by dividing the weights of outgoing arcs by the respective out-degrees:

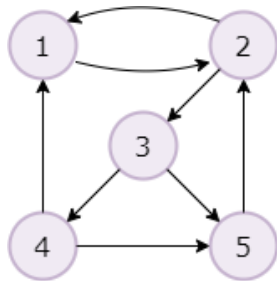
$$A_{norm} = A (D^{out})^{-1} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

$A_{norm}$  is a *column stochastic matrix*.

The modified problem

$$A_{norm} \mathbf{s} = \mathbf{s}$$

always has a solution (**why?**), which is referred to as the *eigenvector centrality score*.



# Eigenvector centrality: pro and contra

The eigenvector centrality scheme is used in many applications.

However, it has several drawbacks:

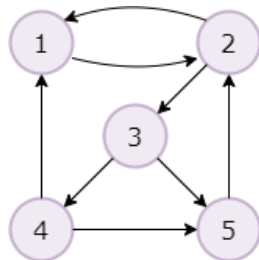
- Can yield multiple scores (= not unique)
- Is not defined if there are "dangling" nodes (i.e., nodes without outgoing connections).

We need a mathematical instrument to analyse such problems in a systematic way.

The main tool is the **Perron-Frobenius th.** and its friends, e.g., **Gershgorin's circle th.**

⋮

Eigenvector centrality has a unique and well defined solution if the underlying directed graph is **connected**.



# Eigenvector centrality: pro and contra

The eigenvector centrality scheme is used in many applications.

However, it has several drawbacks:

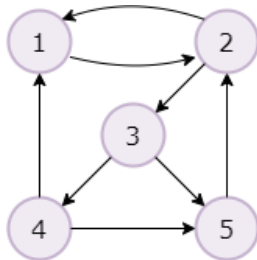
- Can yield multiple scores (= not unique)
- Is not defined if there are "dangling" nodes (i.e., nodes without outgoing connections).

We need a mathematical instrument to analyse such problems in a systematic way.

The main tool is the **Perron-Frobenius th.** and its friends, e.g., **Gershgorin's circle th.**

⋮

Eigenvector centrality has a unique and well defined solution if the underlying directed graph is **connected**.



# Further improvement: PageRank

The problem with EC is that the matrix  $A_{norm}$  is only *non-negative*. Hence, it has to satisfy some additional properties (be irreducible  $\Leftrightarrow$  graph be connected).

Let's approach this problem from a different side and make the matrix **positive!** Define

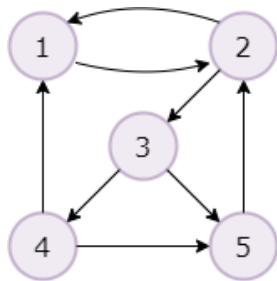
$$A_{PR} = \alpha A_{norm} + (1 - \alpha) \frac{1}{n} J,$$

where  $n = |V|$ ,  $J = \mathbf{1} \cdot \mathbf{1}^T$ , and  $\alpha \in (0, 1)$  ( $\alpha = 0.85$ ).

The **PageRank score** is defined as the solution to

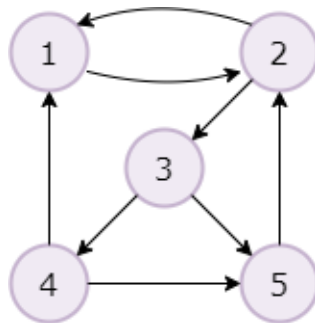
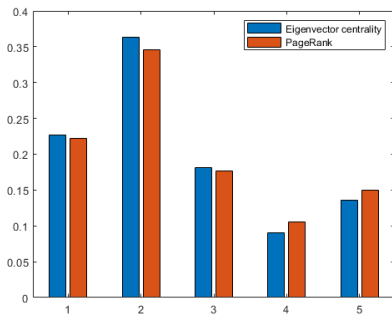
$$A_{PR} s = s.$$

This solution always exists and is unique.





# Comparison: PageRank vs. Eigenvector centrality



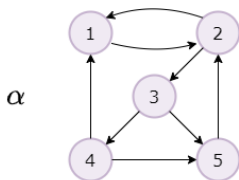
# Interpretation: PageRank

The PageRank algorithm can be neatly interpreted in terms of a Markov chain. Consider the transition matrix

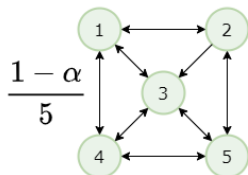
$$A_{PR} = \alpha A_{norm} + (1 - \alpha) \frac{1}{n} J,$$

The elements of  $A_{norm}$  describe the probability of moving from node  $i$  to node  $j \in OUT(i)$ , while the second term describes the probabilities of a random transition from node  $i$  to a random node with equal probability.

The PageRank score is thus the *stationary distribution* of the considered MC.



+



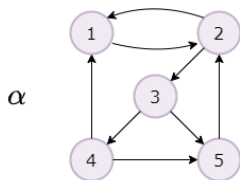
# Interpretation: PageRank

The PageRank algorithm can be neatly interpreted in terms of a Markov chain. Consider the transition matrix

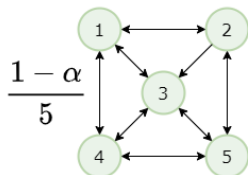
$$A_{PR} = \alpha A_{norm} + (1 - \alpha) \frac{1}{n} J,$$

The elements of  $A_{norm}$  describe the probability of moving from node  $i$  to node  $j \in OUT(i)$ , while the second term describes the probabilities of a random transition from node  $i$  to a random node with equal probability.

The PageRank score is thus the *stationary distribution* of the considered MC.



+





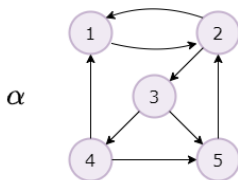
# Interpretation: PageRank

The PageRank algorithm can be neatly interpreted in terms of a Markov chain. Consider the transition matrix

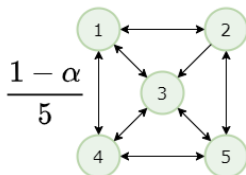
$$A_{PR} = \alpha A_{norm} + (1 - \alpha) \frac{1}{n} \mathbf{J},$$

The elements of  $A_{norm}$  describe the probability of moving from node  $i$  to node  $j \in OUT(i)$ , while the second term describes the probabilities of a random transition from node  $i$  to a random node with equal probability.

The PageRank score is thus the *stationary distribution* of the considered MC.



+



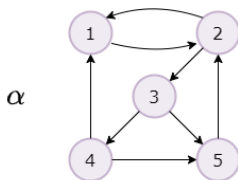
# Interpretation: PageRank

The PageRank algorithm can be neatly interpreted in terms of a Markov chain. Consider the transition matrix

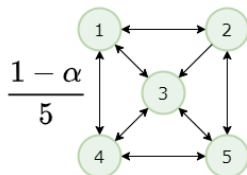
$$A_{PR} = \alpha A_{norm} + (1 - \alpha) \frac{1}{n} J,$$

The elements of  $A_{norm}$  describe the probability of moving from node  $i$  to node  $j \in OUT(i)$ , while the second term describes the probabilities of a random transition from node  $i$  to a random node with equal probability.

The PageRank score is thus the *stationary distribution* of the considered MC.



+



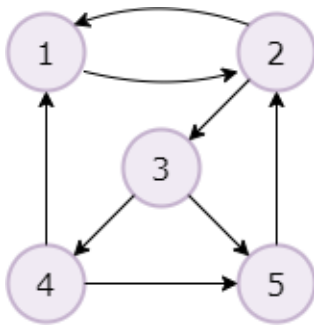
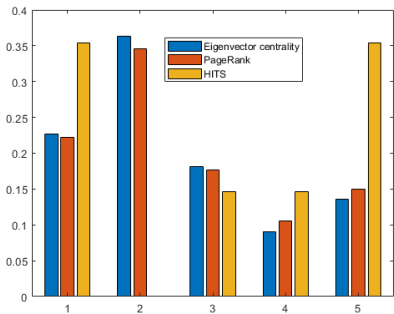








# PageRank vs. Eigenvector centrality vs. HITS















# Quasi exponential ranking


The main drawback of exponential ranking is that it is non-linear.

Therefore, it was suggested<sup>1</sup> to use a (partially) linearized scheme (EXP), which we call *quasi exponential ranking*.

$$\mathbf{1}^\top \left[ \mathbf{1} + \frac{1}{\mu} \mathbf{A} \mathbf{p} \right] \mathbf{p} = \mathbf{1} + \frac{1}{\mu} \mathbf{A} \mathbf{p}. \quad (\text{qEXP})$$

This scheme has a unique solution  $\mathbf{p}^*$  if  $\mu > \max_{a_{ij} \leq 0} |(a_{ij})|$  (which agrees with (EXP)). Furthermore,  $\mathbf{p}^*$  is the eigenvector of  $\frac{1}{\mu} \mathbf{A} + \mathbf{J}$  corresponding to the spectral radius  $\rho \left( \frac{1}{\mu} \mathbf{A} + \mathbf{J} \right)$ .

---

<sup>1</sup>Gromov D., Evmenova E. On the Exponential Ranking and Its Linear Counterpart (2022) Studies in Computational Intelligence, 1015, pp. 260 - 270. 

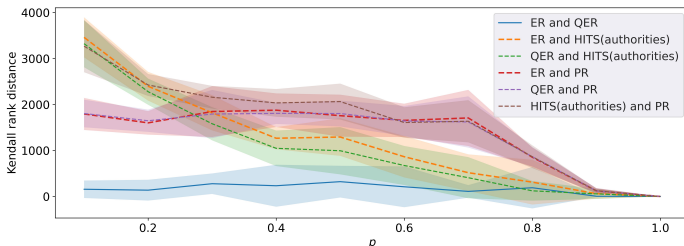




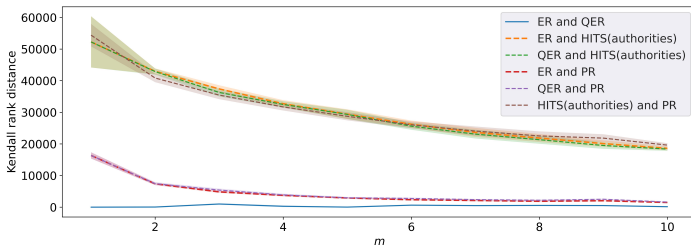


# Quasi exponential ranking: numerical comparison

Erdős-Rényi graph with 500 nodes and parameter  $p \in \{0.1, 0.2, \dots, 1\}$ :



Barabási-Albert graph with 500 nodes and parameter  $m \in \{1, 2, \dots, 10\}$ :





# Quasi exponential ranking: numerical comparison

R-ary Tree graph with 500 nodes and parameter  $r \in \{1, 2, \dots, 10\}$ :

