PageRank: Ranking chemicals based on heterogeneous data

Amelie Klein and Uziel Gonzalez

Amelie Klein and Uziel Gonzalez

Introduction

Google's PageRank method was developed to evaluate the importance of web-pages via their link structure. The mathematics of PageRank, however, are entirely general and apply to any graph or network in any domain.

It's even used for systems analysis of road networks, as well as biology, chemistry, neuroscience, and physics.

Data

When you are seeking a herbicidal, many chemicals should be tested. Those can be applied to multiple plants of multiple species at various doses and measuring the responses in many screens.

Data

Compound	Rate	Method	Specie 1	Specie 2	Specie 3	Specie 4	Specie 5	Specie 6	Specie 7
1	1000	EPS POST	60	90	70	90	0	70	100
2	1000	EPS POST	0	30	0	0	0	0	0
3	1000	EPS POST	100	100	80	100	0	90	100
3	1000	EPS POST	100	100	90	100	0	90	100
4	1000	EPS POST	0	30	0	10	0	0	30
4	1000	EPS POST	20	100	20	60	0	10	70
5	1000	EPS POST	0	60	20	30	0	10	50
6	1000	EPS POST	20	60	30	30	0	20	60
7	1000	EPS POST	40	80	90	40	0	90	90
7	1000	EPS POST	20	80	90	70	0	80	80
8	1000	EPS POST	90	100	90	100	0	90	90
9	1000	EPS POST	10	90	80	100	0	80	90
10	1000	EPS POST	90	80	90	80	0	90	90

Toy example

Transition probability

Dosis	Compound	Specie1	Specie2	Specie3
50	1	60	90	0
50	2	30	10	20
50	3	50	40	70

0	20	70
60+90	0	50+40+50
60+90	0	0

From our graph, we construct a transition matrix, and calculate its left eigenvector.

This eigenvector gives us a score for each chemical.

This scores give us a ranking,

Analysis of Data

Compound	Rate	Method	Specie 1	Specie 2	Specie 3	Specie 4	Specie 5	Specie 6	Specie 7
1	1000	EPS POST	60	90	70	90	0	70	100
2	1000	EPS POST	0	30	0	0	0	0	0
3	1000	EPS POST	100	100	80	100	0	90	100
3	1000	EPS POST	100	100	90	100	0	90	100
4	1000	EPS POST	0	30	0	10	0	0	30
4	1000	EPS POST	20	100	20	60	0	10	70
5	1000	EPS POST	0	60	20	30	0	10	50
6	1000	EPS POST	20	60	30	30	0	20	60
7	1000	EPS POST	40	80	90	40	0	90	90
7	1000	EPS POST	20	80	90	70	0	80	80
8	1000	EPS POST	90	100	90	100	0	90	90
9	1000	EPS POST	10	90	80	100	0	80	90
10	1000	EPS POST	90	80	90	80	0	90	90

A first approach
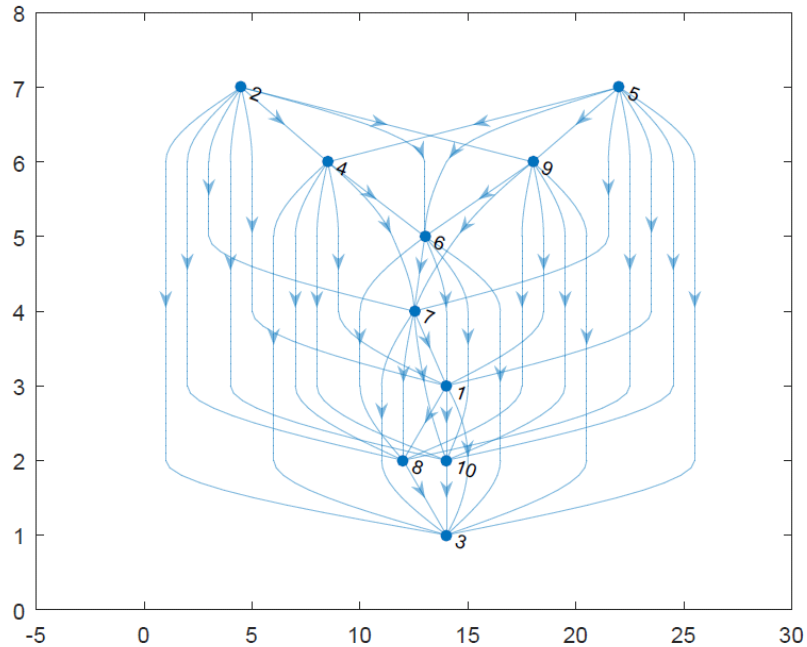
1] 0.1787
2] 0.0868
3] 0.8669
4] 0.0910
5] 0.0868
6] 0.1046
7] 0.1220
8] 0.2820
9] 0.0910
10] 0.2820

Future work

References

PageRank beyond the Web by
David F. Gleich
2015 Society for Industrial and Applied Mathematics

Graph Theory and Linear Algebra of Google's Pagerank
(http://chipjacks.com)

The predictive power of ranking systems in association football
Jan Lasek