MODELING CITATION FLOW - CAN WE TRUST JOURNAL RANKINGS?

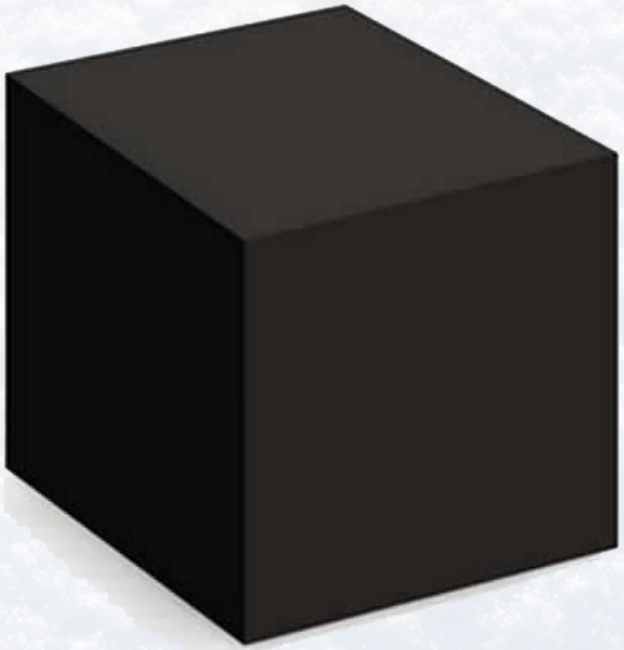
Ludvig Bohlin

NetSci 2014


UMEÅ UNIVERSITET

PROBLEM

The black box of journal rankings



HOW TO EVALUATE SCIENCE?

- **Citation Network**

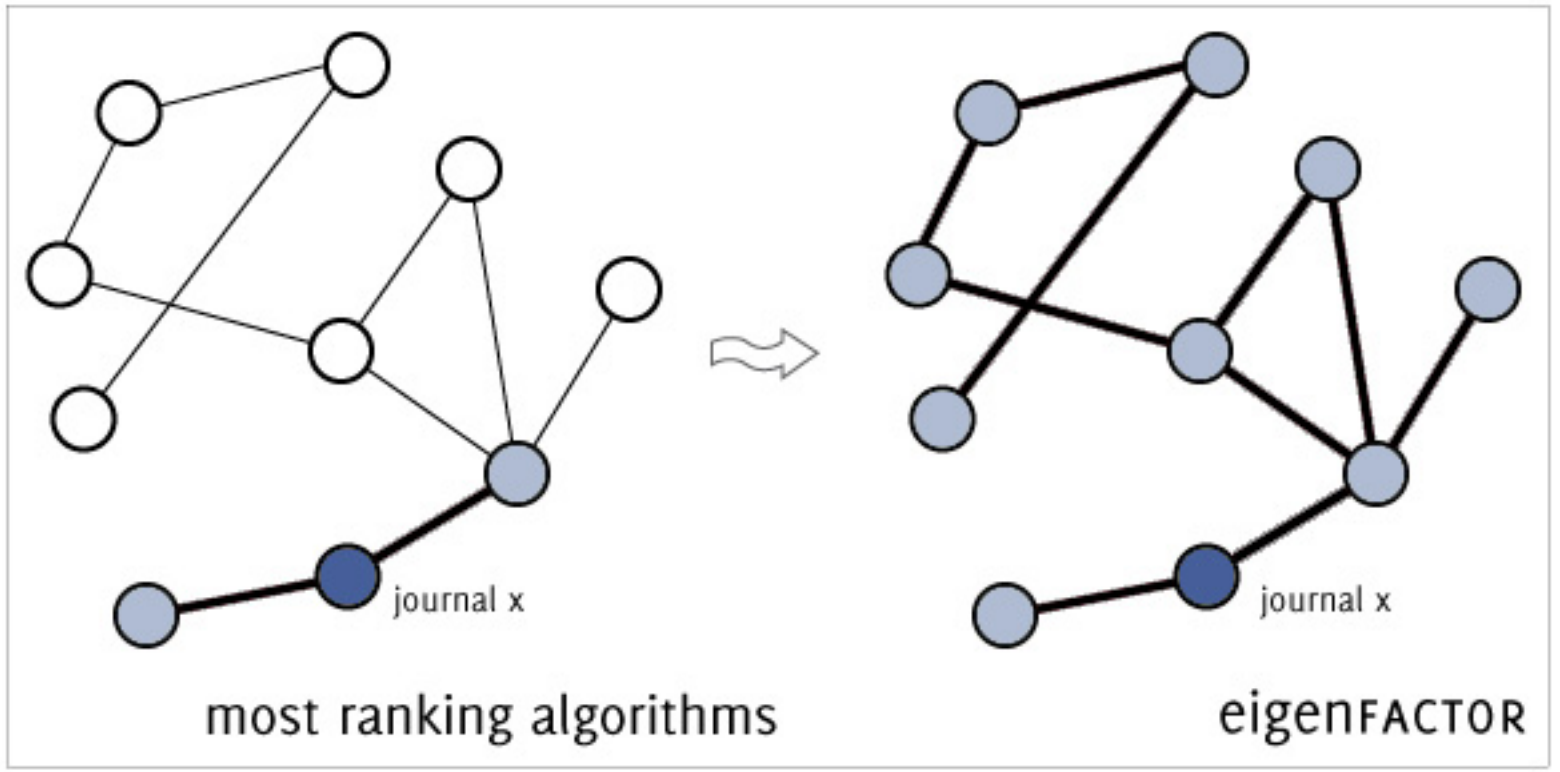
Metrics can be seen as flow models of different Markov order that seek to capture researchers navigating scholarly literature

IMPACT FACTOR

- *A* = Citations in 2006 and 2007 from articles in indexed journals during 2008
- *B* = the total number of "citable items" published by journal in 2006 and 2007
- Impact factor 2008 = *A*/*B*

Zero-order model: Independent of the currently visited journal

EIGENFACTOR



First-order model: Depends only on the currently visited journal

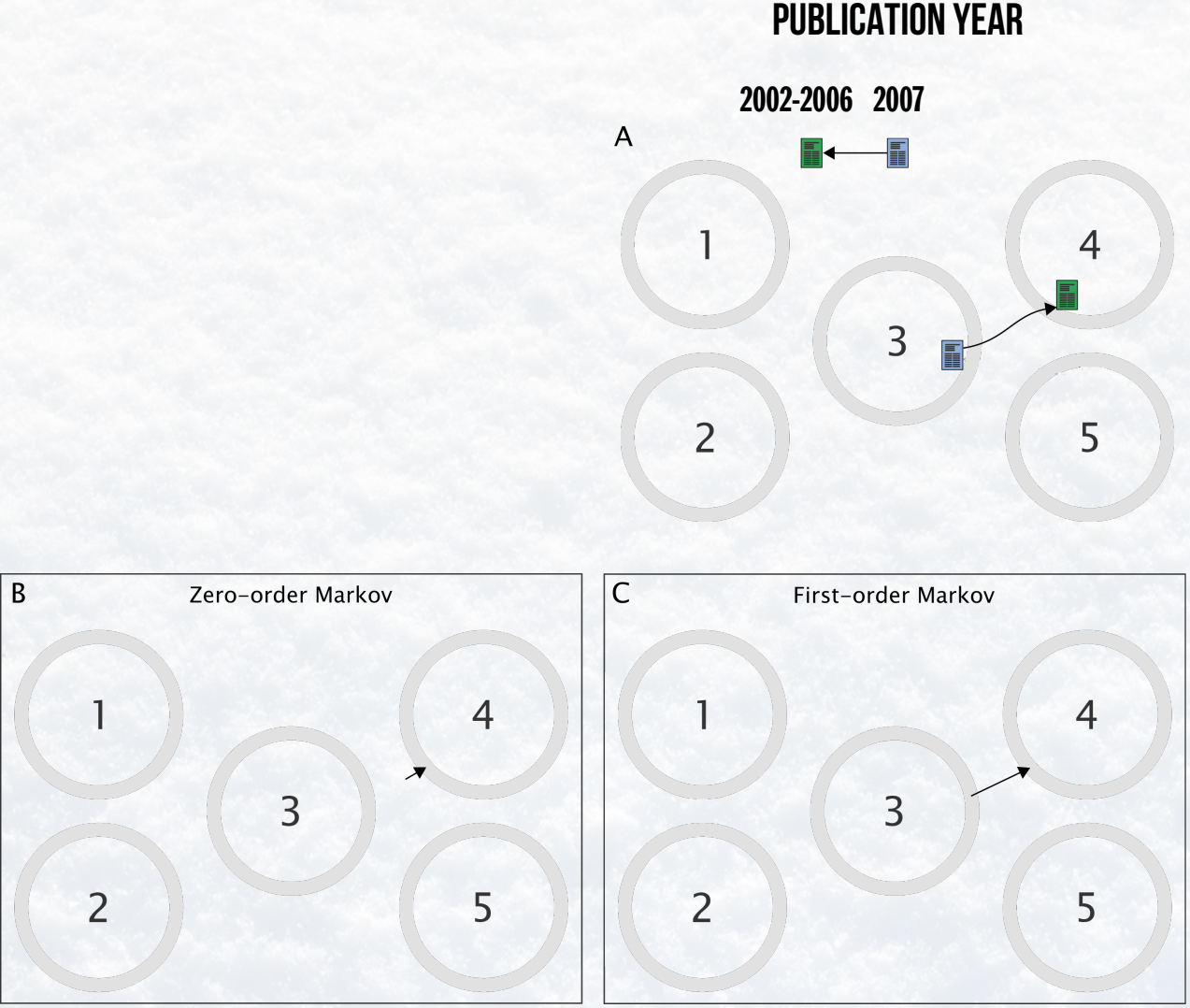
SECOND-ORDER MODEL?

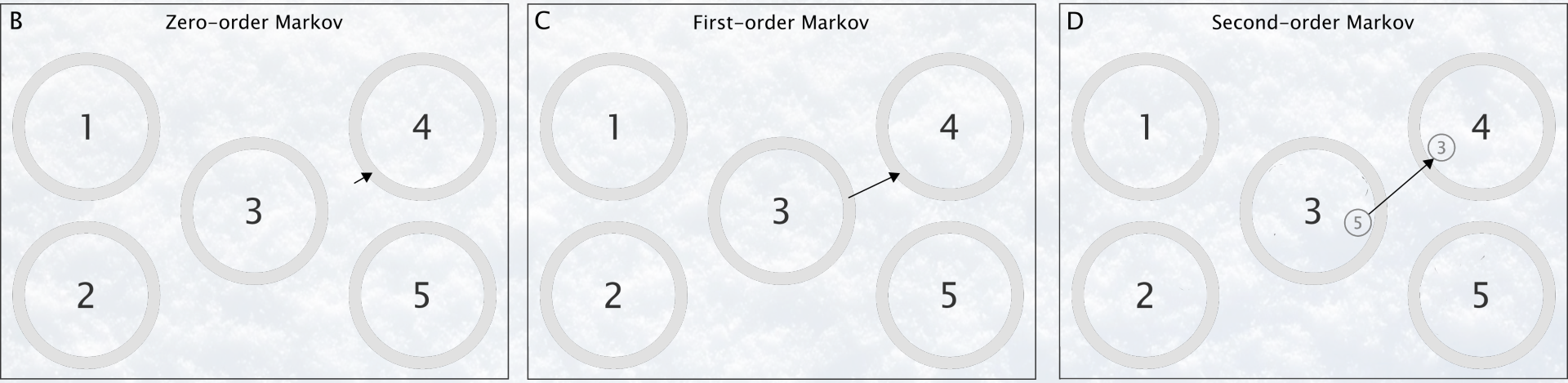
Second-order model: Depend both on the currently visited journal *and* the previously visited journal.

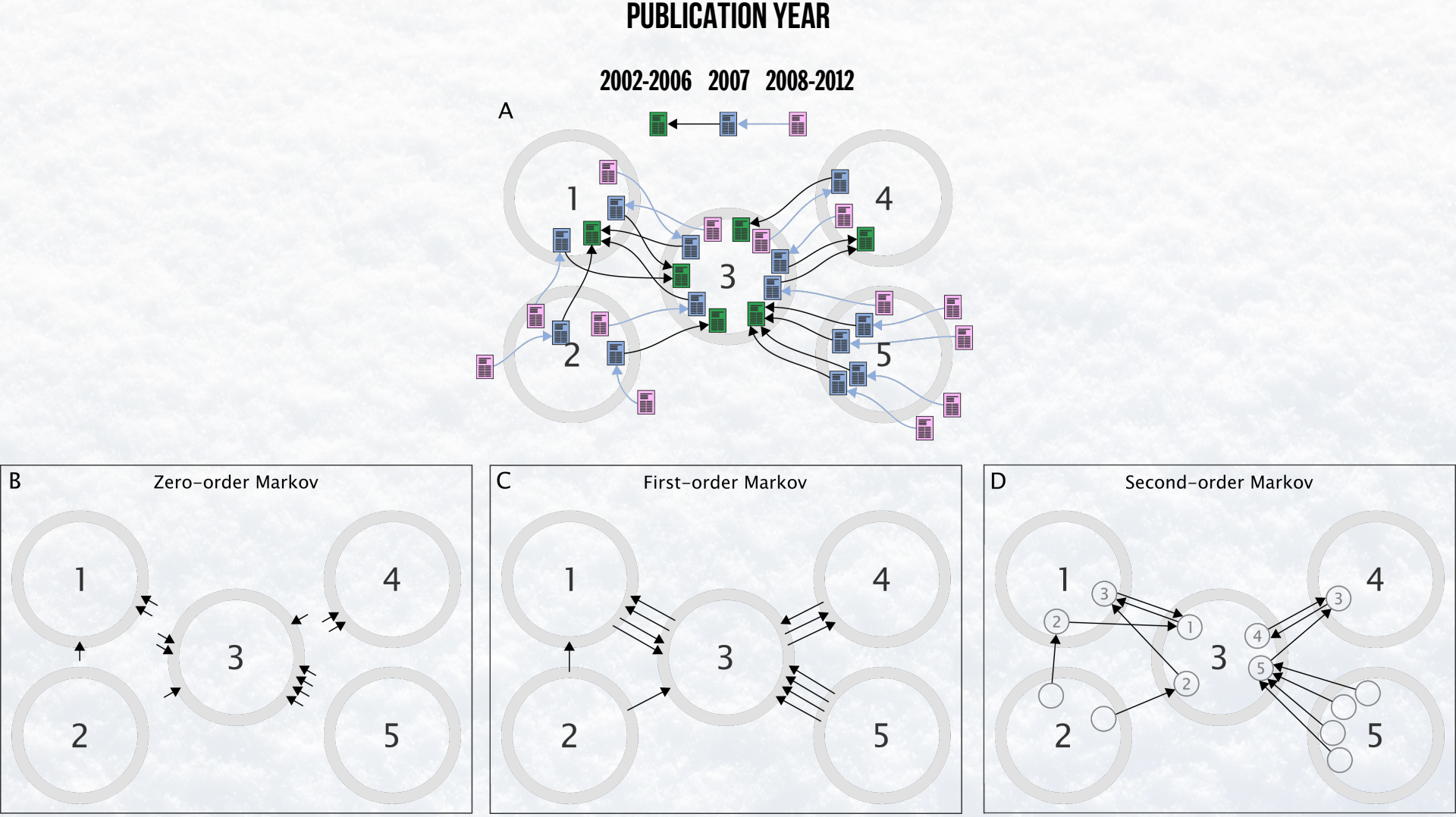
CAN WE TRUST JOURNAL RANKINGS?

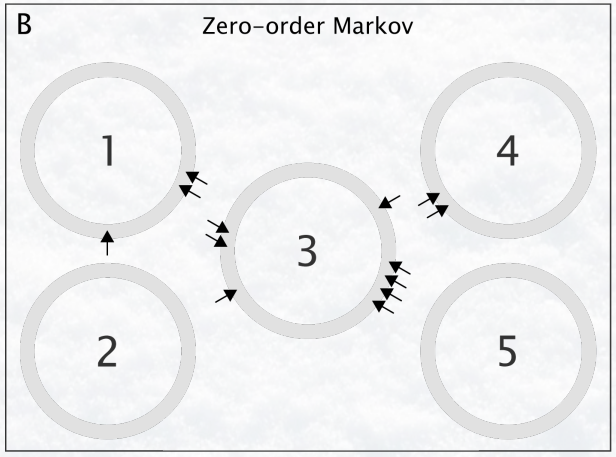
DATA

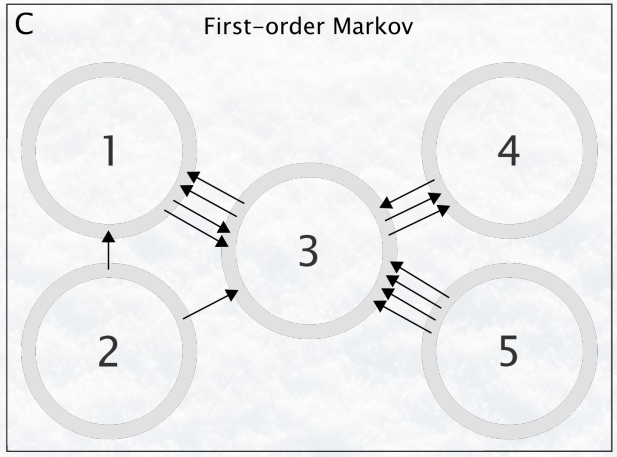
- Thomson Reuters Web of Science
- Year 2007

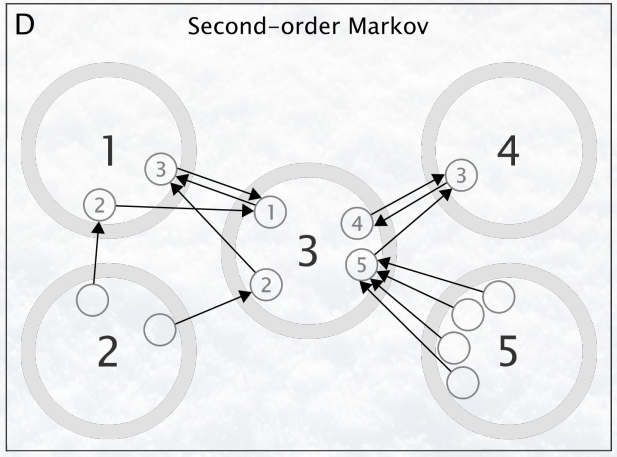

PUBLICATION YEAR
2002-2006
2007
A
1
2
3
4
5
B
Zero-order Markov
1
2
3
4
5
C
First-order Markov
1
2
3
4
5

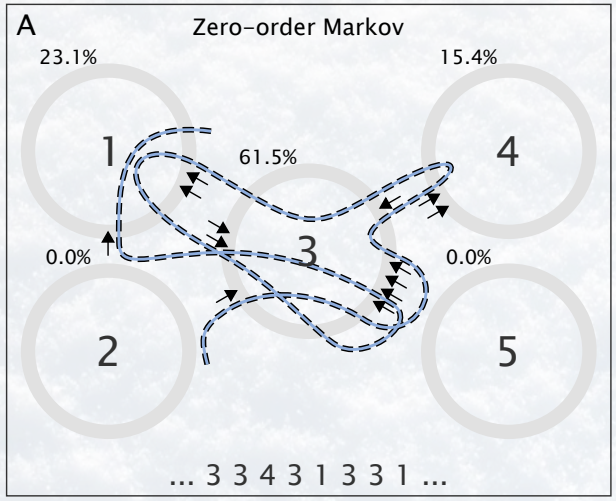

PUBLICATION YEAR
2002-2006
2007
2008-2012
A
1
2
3
4
5
B
Zero-order Markov
C
First-order Markov
D
Second-order Markov

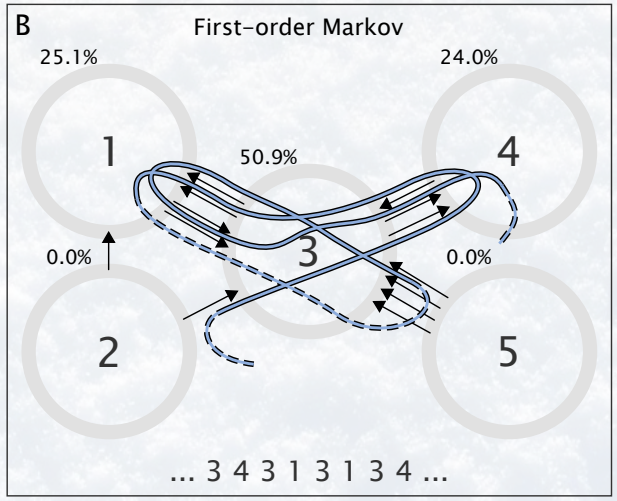

PUBLICATION YEAR
2002-2006
2007
2008-2012
A
1
2
3
4
5
B
Zero-order Markov
1
2
3
4
5
C
First-order Markov
1
2
3
4
5
D
Second-order Markov
1
2
3
4
5

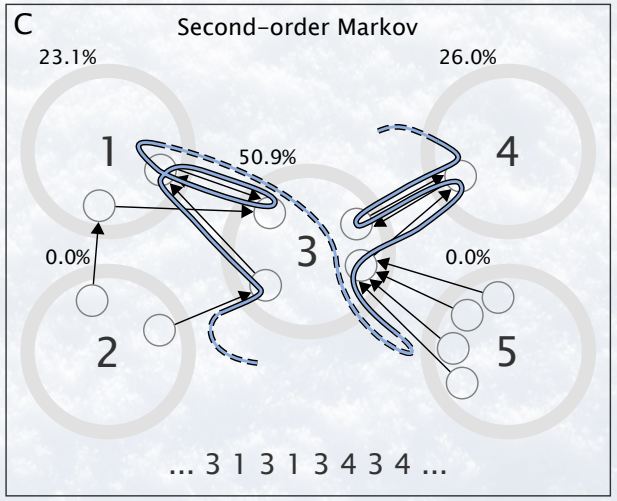

B
Zero-order Markov
1
2
3
4
5


C
First-order Markov
1
2
3
4
5

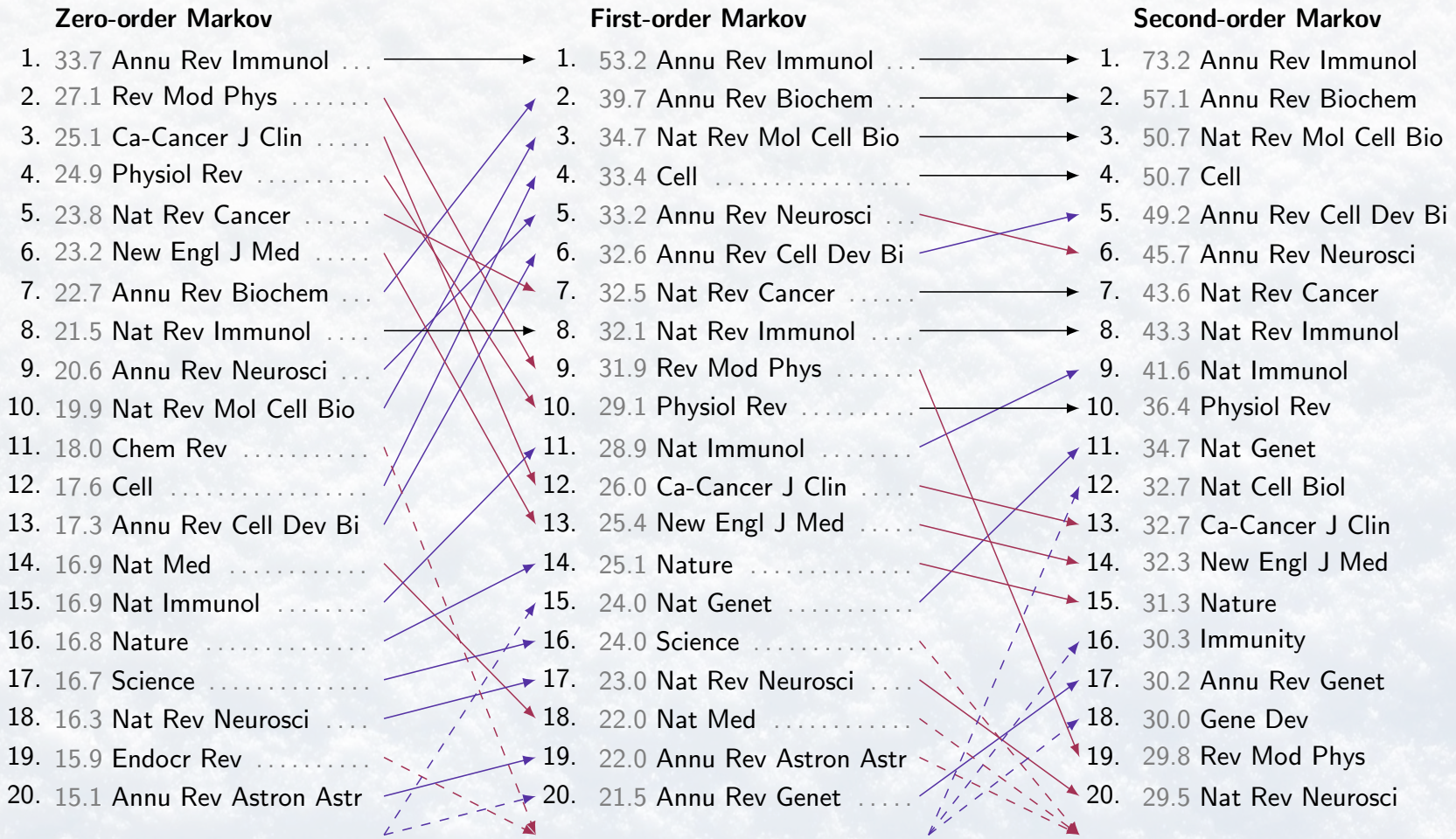

D
Second-order Markov
1
2
3
4
5


A
Zero-order Markov
23.1%
15.4%
61.5%
0.0%
0.0%
1
2
3
4
5
... 3 3 4 3 1 3 3 1 ...

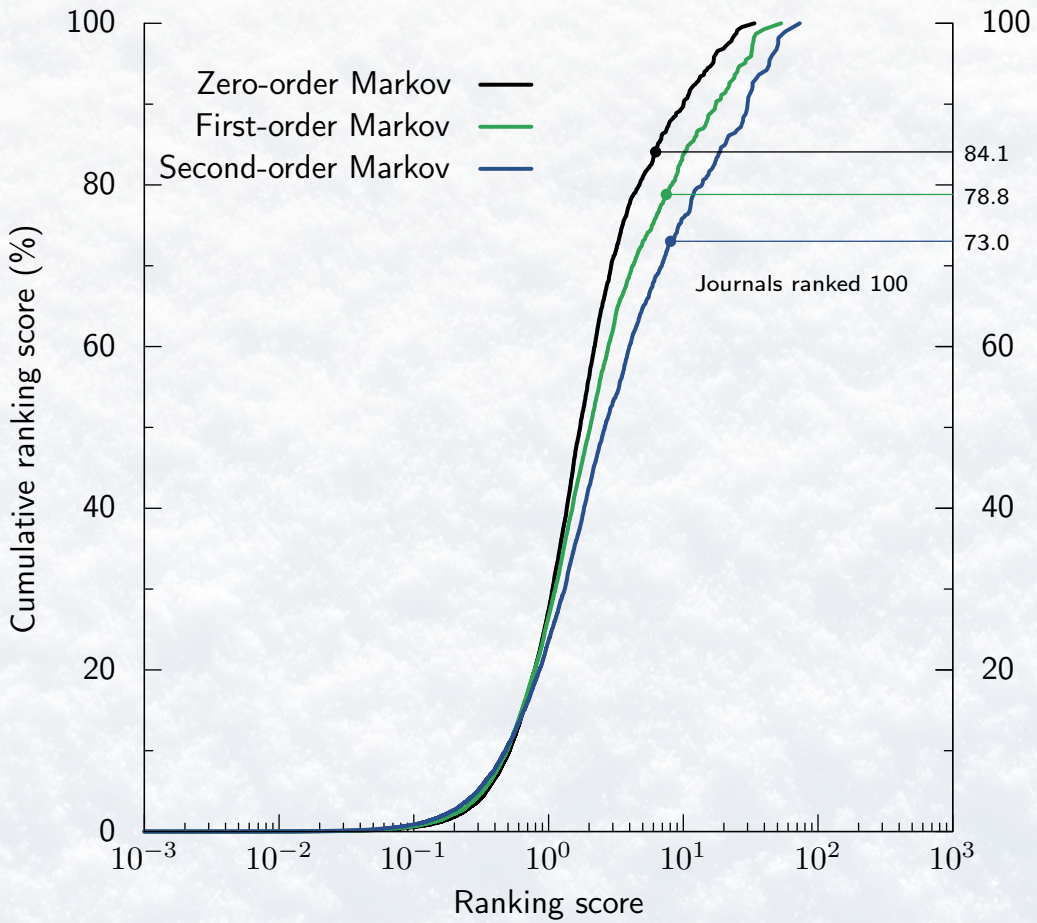

B
First-order Markov
25.1%
24.0%
50.9%
0.0%
0.0%
1
2
3
4
5
... 3 4 3 1 3 1 3 4 ...


C
Second-order Markov
23.1%
26.0%
50.9%
0.0%
0.0%
1
2
3
4
5
... 3 1 3 1 3 4 3 4 ...

RESULTS


Zero-order Markov
1. 33.7 Annu Rev Immunol . . .
2. 27.1 Rev Mod Phys . . . . . .
3. 25.1 Ca-Cancer J Clin . . . . .
4. 24.9 Physiol Rev . . . . . .
5. 23.8 Nat Rev Cancer . . . . . .
6. 23.2 New Engl J Med . . . . .
7. 22.7 Annu Rev Biochem . . .
8. 21.5 Nat Rev Immunol . . . .
9. 20.6 Annu Rev Neurosci . . .
10. 19.9 Nat Rev Mol Cell Bio
11. 18.0 Chem Rev . . . . . .
12. 17.6 Cell . . . . . .
13. 17.3 Annu Rev Cell Dev Bi
14. 16.9 Nat Med . . . . . .
15. 16.9 Nat Immunol . . . . . .
16. 16.8 Nature . . . . . .
17. 16.7 Science . . . . . .
18. 16.3 Nat Rev Neurosci . . . .
19. 15.9 Endocr Rev . . . . . .
20. 15.1 Annu Rev Astron Astr
First-order Markov
1. 53.2 Annu Rev Immunol . . .
2. 39.7 Annu Rev Biochem . . .
3. 34.7 Nat Rev Mol Cell Bio
4. 33.4 Cell . . . . . .
5. 33.2 Annu Rev Neurosci . . .
6. 32.6 Annu Rev Cell Dev Bi
7. 32.5 Nat Rev Cancer . . . . . .
8. 32.1 Nat Rev Immunol . . . .
9. 31.9 Rev Mod Phys . . . . . .
10. 29.1 Physiol Rev . . . . . .
11. 28.9 Nat Immunol . . . . . .
12. 26.0 Ca-Cancer J Clin . . . . .
13. 25.4 New Engl J Med . . . . .
14. 25.1 Nature . . . . . .
15. 24.0 Nat Genet . . . . . .
16. 24.0 Science . . . . . .
17. 23.0 Nat Rev Neurosci . . . .
18. 22.0 Nat Med . . . . . .
19. 22.0 Annu Rev Astron Astr
20. 21.5 Annu Rev Genet . . . . .
Second-order Markov
1. 73.2 Annu Rev Immunol
2. 57.1 Annu Rev Biochem
3. 50.7 Nat Rev Mol Cell Bio
4. 50.7 Cell
5. 49.2 Annu Rev Cell Dev Bi
6. 45.7 Annu Rev Neurosci
7. 43.6 Nat Rev Cancer
8. 43.3 Nat Rev Immunol
9. 41.6 Nat Immunol
10. 36.4 Physiol Rev
11. 34.7 Nat Genet
12. 32.7 Nat Cell Biol
13. 32.7 Ca-Cancer J Clin
14. 32.3 New Engl J Med
15. 31.3 Nature
16. 30.3 Immunity
17. 30.2 Annu Rev Genet
18. 30.0 Gene Dev
19. 29.8 Rev Mod Phys
20. 29.5 Nat Rev Neurosci

RESULTS



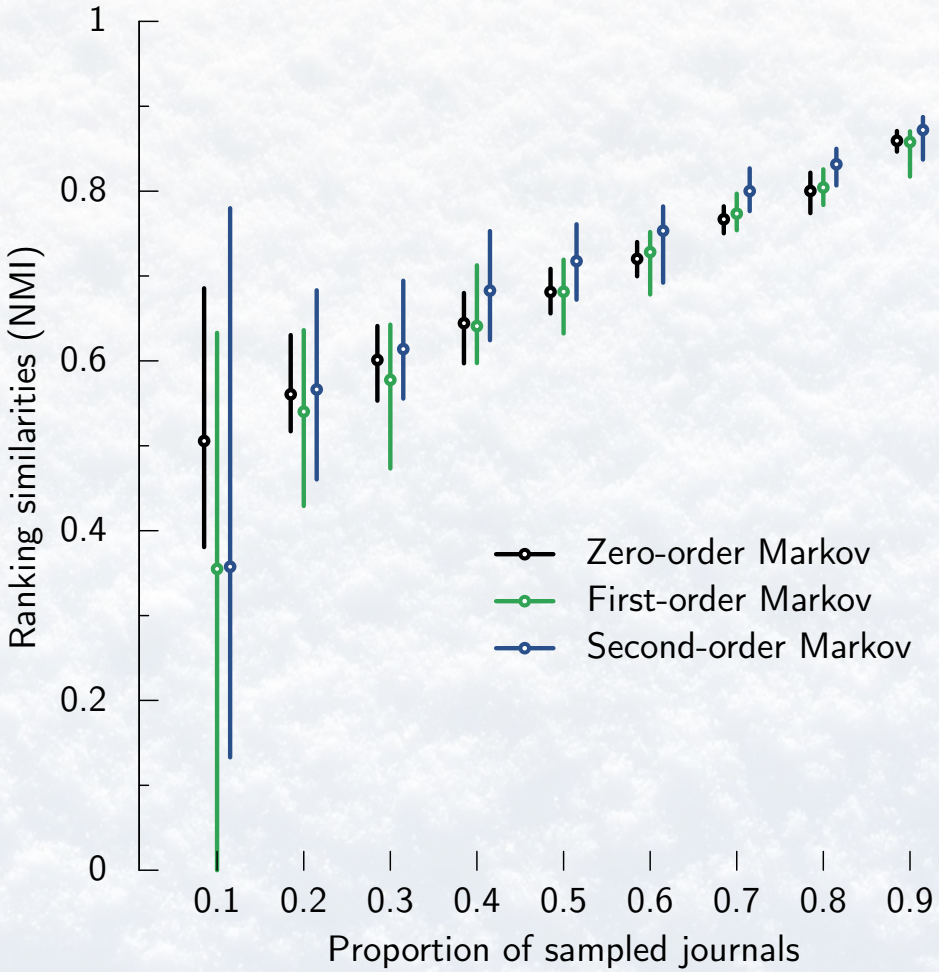
Top 100 journals

16 %

21 %

27 %

RESULTS


1
0.8
0.6
0.4
0.2
0
Ranking similarities (NMI)
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Proportion of sampled journals
Zero-order Markov
First-order Markov
Second-order Markov

SUMMARY

	Local noise	Flow propagation	Time proximity
0th order	✗	✗	✓
1st order	✓	✗	✓
2nd order	✓	✓	✗

Physics > Physics and Society

Robustness of journal rankings by network flows with different amounts of memory

Ludvig Bohlin, Alcides Viamontes Esquivel, Andrea Lancichinetti, Martin Rosvall

(Submitted on 30 May 2014)

As the number of scientific journals has multiplied, journal rankings have become increasingly important for scientific decisions. From submissions and subscriptions to grants and hirings, researchers, policy makers, and funding agencies make important decisions with influence from journal rankings such as the ISI journal impact factor. Typically, the rankings are derived from the citation network between a selection of journals and unavoidably depend on this selection. However, little is known about how robust rankings are to the selection of included journals. Here we compare the robustness of three journal rankings based on network flows induced on citation networks. They model pathways of researchers navigating scholarly literature, stepping between journals and remembering their previous steps to different degree: zero-step memory as impact factor, one-step memory as Eigenfactor, and two-step memory, corresponding to zero-, first-, and second-order Markov models of citation flow between journals. We conclude that a second-order Markov model is slightly more robust, because it combines the advantages of the lower-order models: perturbations that remain local and citation weights that depend on journal importance. However, the robustness gain comes at the cost of requiring more data, because the second-order Markov model requires citation data from twice as long a period.

THANKS FOR LISTENING!