CWTS
Meaningful metrics

Uncovering important intermediate publications

L. Šubelj[1], L. Waltman[2], V.A. Traag[2], N.J. van Eck[2]

[1]Faculty of Computer and Information Science, University of Ljubljana, Slovenia

[2]Centre for Science and Technology Studies, Leiden University, the Netherlands

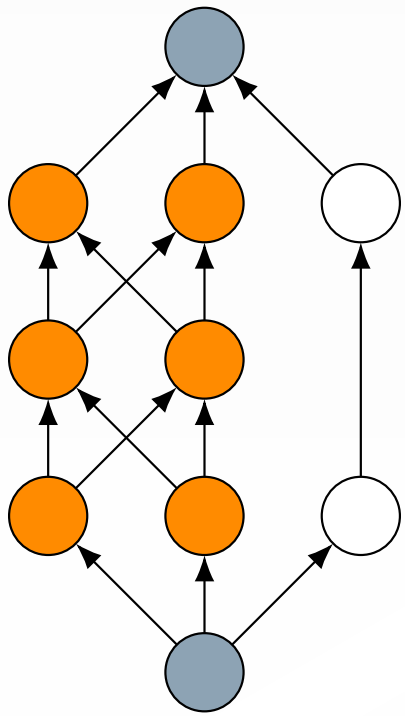
30 June 2018

Sunbelt, Utrecht, the Netherlands

Universiteit Leiden

Intermediate publications

- Historiography: describe development of a field.
- What publications have been important in that development?
- Rely on citations to identify important intermediate publications.



CWTS
Meaningful metrics

Existing technique

Main path analysis

- Relies on traversal counts of edges.
- Selects path(s) that have a high sum of traversal counts.
- Rewards relatively long paths.
- Conceptually unclear, not always clear results.

Shortest or longest paths

- Shortest paths typically do not include most important publications.
- Longest paths typically include many irrelevant publications.

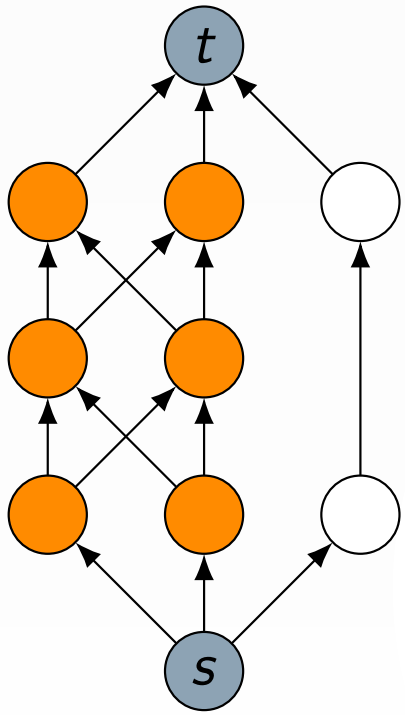
CWTS
Meaningful metrics

Main idea

Important intermediate publications should be well connected.

Illustration

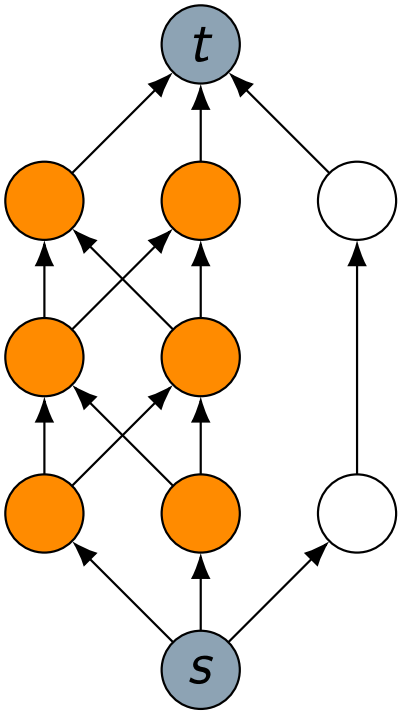
Multiple paths



Short paths

CWTS
Meaningful metrics

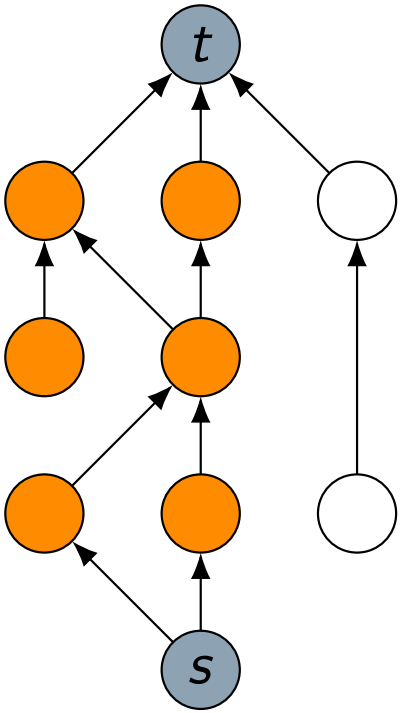
Illustration



- Only some references are relevant.
- Reference is relevant with probability p.
- Is there a path (of relevant references) through a node?
- This is *intermediacy*.

CWTS
Meaningful metrics

Illustration



- Only some references are relevant.
- Reference is relevant with probability p.
- Is there a path (of relevant references) through a node?
- This is *intermediacy*.

CWTS
Meaningful metrics

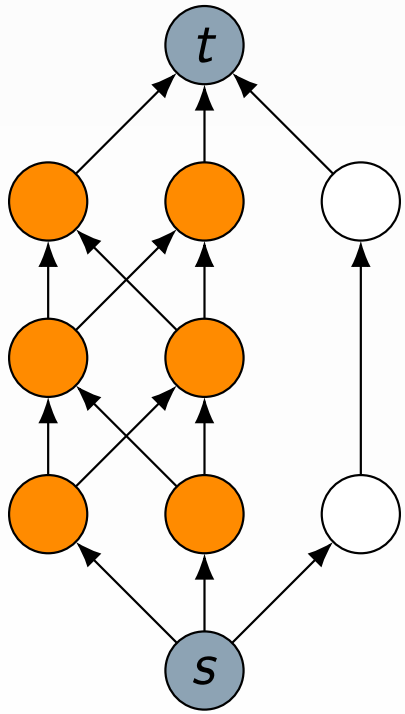
Formal notation

- The probability there is a path from i to j is $\Pr(X_{ij})$.
- Intermediacy is the probability node u lies on a path from s to t.
- Intermediacy of node u from source s to target t is

$$\phi_u = \Pr(X_{st}^u) = \Pr(X_{su}) \Pr(X_{ut})$$

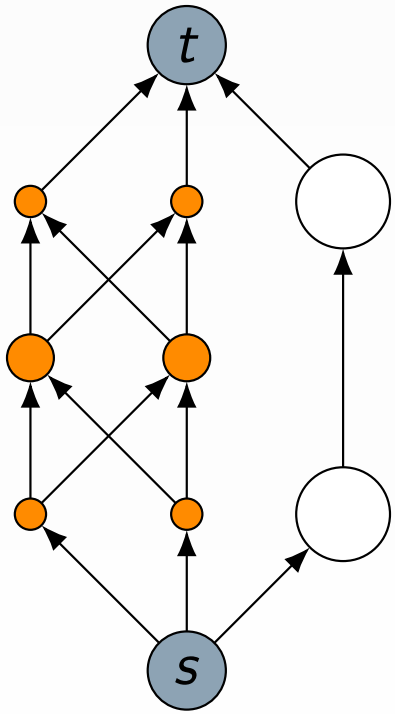
CWTS
Meaningful metrics

How does intermediacy behave?



Behaviour

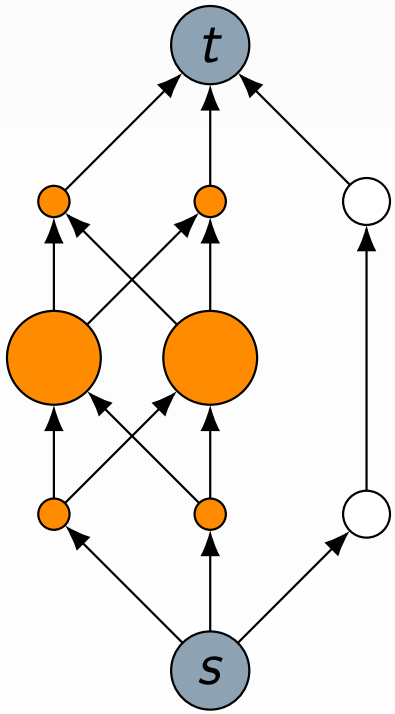
For $p \to 0$ shortest paths are most important.



CWTS
Meaningful metrics

Behaviour

For $p \to 1$ number of independent paths are most important.

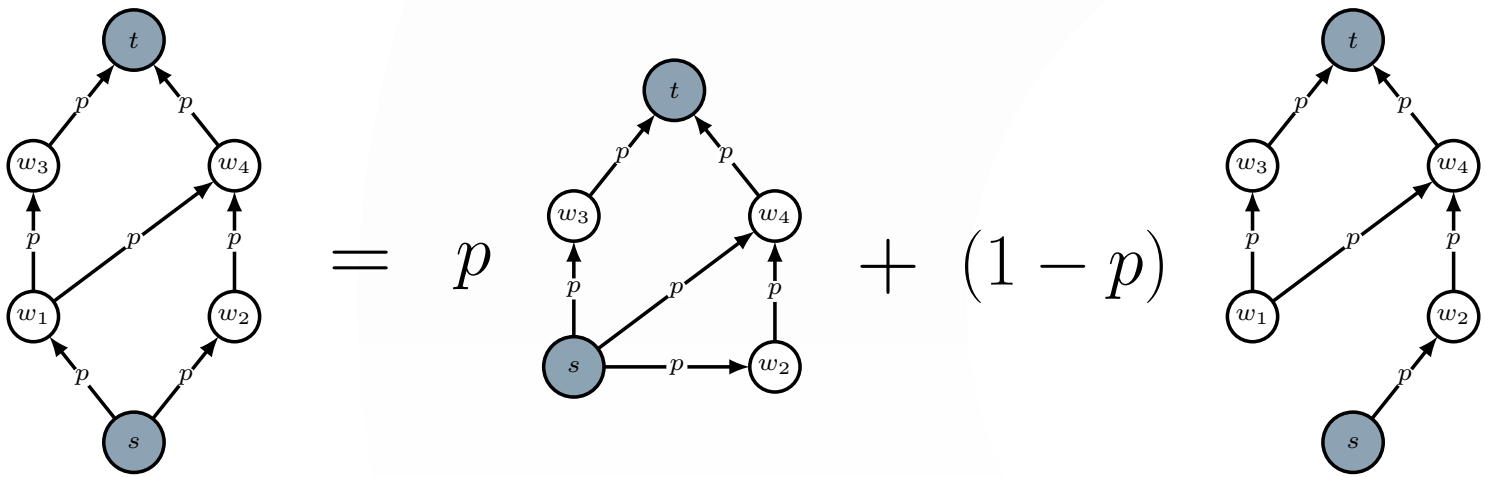


CWTS
Meaningful metrics

Exact algorithm

Decomposition algorithm by edge contraction & removal

$$\Pr(X_{st} \mid G) = p \Pr(X_{st} \mid G/e) + (1-p) \Pr(X_{st} \mid G-e)$$



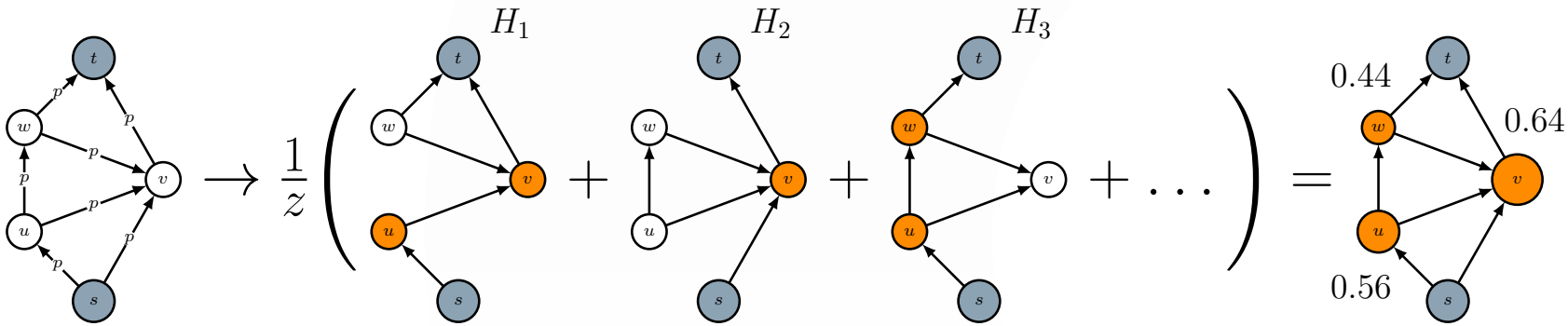
Runs in exponential time

CWTS
Meaningful metrics

Approximate algorithm

Simple Monte Carlo simulation algorithm by sampling

$$\phi_u = \Pr(X^u_{st} \mid G) = \frac{1}{z} \sum_{k=1}^{z} \mathrm{I}(X^u_{st} \mid H_k)$$



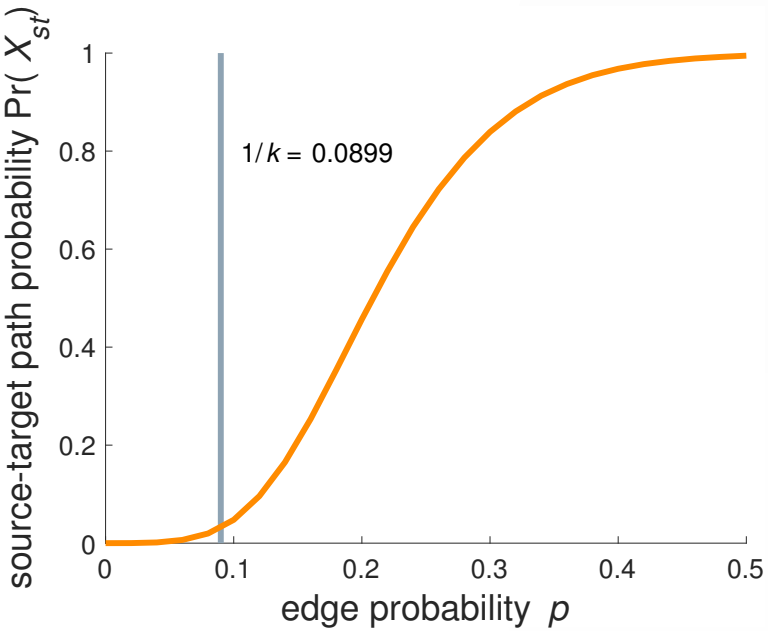
Runs in linear time using probabilistic depth-first search.

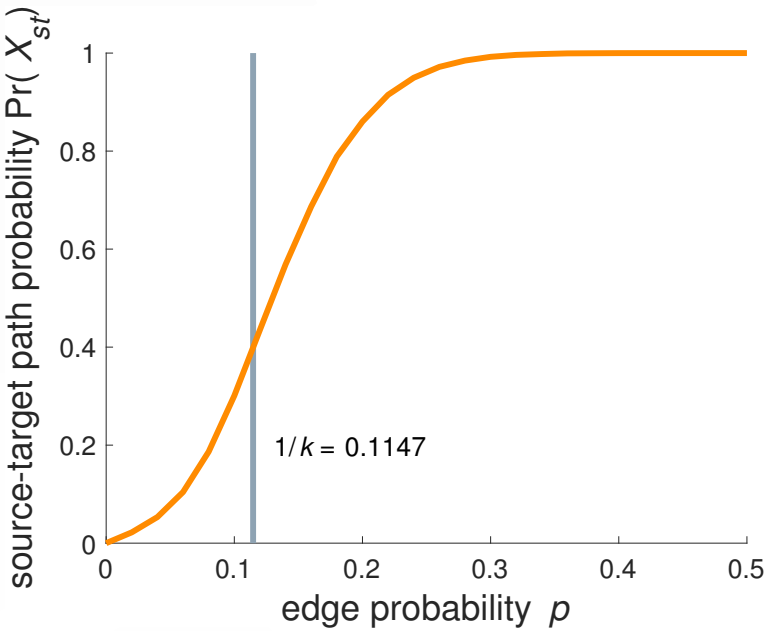
CWTS
Meaningful metrics

Phase transition

For what p does s–t path exist and is intermediacy $\phi_u > 0$?

$$p \geq n/2m = 1/k$$

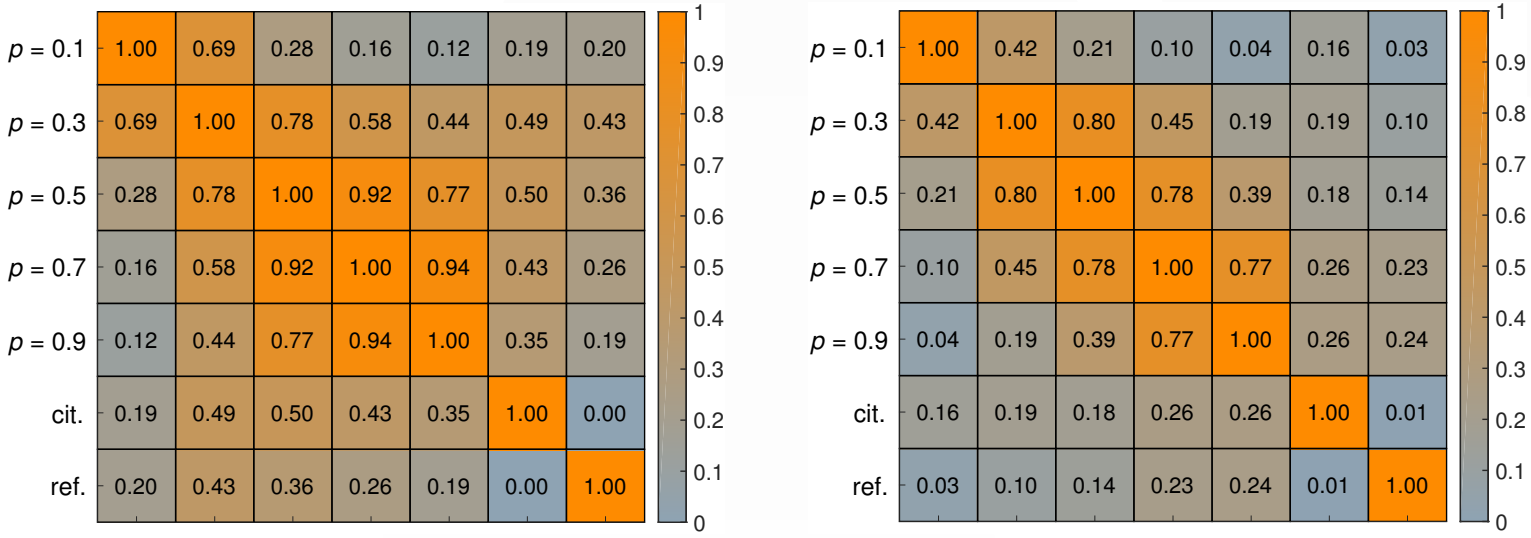




CWTS
Meaningful metrics

Intermediacy $\neq$ centrality

Correlation between intermediacy and citations/references



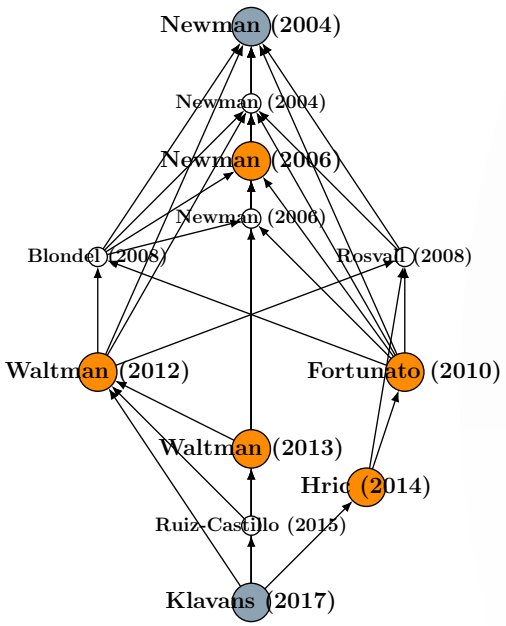
Intermediacy ϕ uncorrelated with standard centrality measures

CWTS Meaningful metrics

Modularity example

source Klavans & Boyack (2017), Which type of citation analysis generates, *JASIST* **68**(4), 984-998.

target Newman & Girvan (2004), Finding and evaluating community structure in networks, *Phys. Rev. E* **69**(2), 026113.



1 Waltman & Van Eck (2013), A smart local moving algorithm for large-scale modularity-based community detection, *EPJB* **86**, 471.

2 Waltman & Van Eck (2012), A new methodology for constructing a publication-level classification system…, *JASIST* **63**(12), 2378-2392.

3 Hric et al. (2014), Community detection in networks: Structural communities versus ground truth, *Phys. Rev. E* **90**(6), 062805.

4 Fortunato (2010), Community detection in graphs, *Phys. Rep.* **486**(3-5), 75-174.

5 Newman (2006), Modularity and community structure in networks, *PNAS* **103**(23), 8577-8582.

6 Ruiz-Castillo & Waltman (2015), Field-normalized citation impact indicators using algorithmically…, *J. Informetr.* **9**(1), 102-117.

7 Blondel et al. (2008), Fast unfolding of communities in large networks, *J. Stat. Mech.*, P10008.

8 Newman (2006), Finding community structure in networks using the eigenvectors of matrices, *Phys. Rev. E* **74**(3), 036104.

9 Newman (2004), Fast algorithm for detecting community structure in networks, *Phys. Rev. E* **69**(6), 066133.

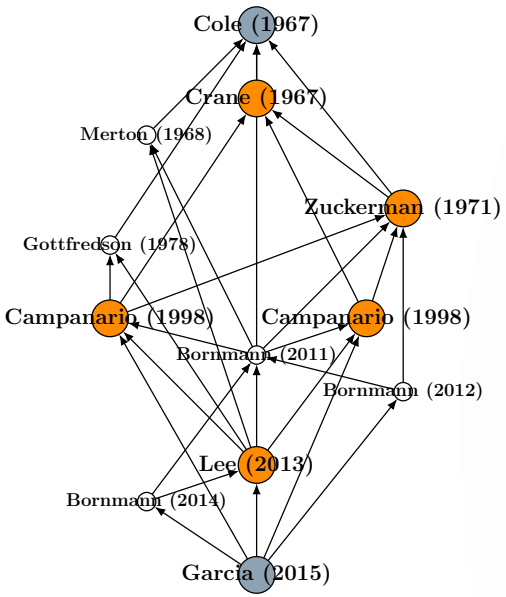
10 Rosvall & Bergstrom (2008), Maps of random walks on complex networks reveal community structure, *PNAS* **105**(4), 1118-1123.

CWTS Meaningful metrics

Peer review example

source Garcia et al. (2015), The author-editor game, *Scientometrics* **104**(1), 361-380.

target Cole & Cole (1967), Scientific output and recognition, *Am. Sociol. Rev.* **32**(3), 377-390.



1 Lee et al. (2013), Bias in peer review, *JASIST* **64**(1), 2-17.

2 Zuckerman & Merton (1971), Patterns of evaluation in science: Institutionalisation, structure and functions..., *Minerva* **9**(1), 66-100.

3 Campanario (1998), Peer review for journals as it stands today: Part 1, *Sci. Commun.* **19**(3), 181-211.

4 Crane (1967), The gatekeepers of science: Some factors affecting the selection of articles for scientific journals, *Am. Sociol.* **2**(4), 195-201.

5 Campanario (1998), Peer review for journals as it stands today: Part 2, *Sci. Commun.* **19**(4), 277-306.

6 Gottfredson (1978), Evaluating psychological research reports: Dimensions, reliability, and correlates..., *Am. Psychol.* **33**(10), 920-934.

7 Bornmann (2011), Scientific peer review, *Annu. Rev. Inform. Sci.* **45**(1), 197-245.

8 Bornmann (2012), The Hawthorne effect in journal peer review, *Scientometrics* **91**(3), 857-862.

9 Bornmann (2014), Do we still need peer review? An argument for change, *JASIST* **65**(1), 209-213.

10 Merton (1968), The Matthew effect in science, *Science* **159**(3810), 56-63.

CWTS Meaningful metrics

Conclusions

Main points

- *Intermediacy* new measure of importance of publications.
- Favours short paths & many independent paths.
- Shows promising results in case studies.

Future work

- Axiomatic framework for path probability.
- Applicability on general (directed) graphs?

CWTS
Meaningful metrics

Paper soon on arXiv.org

Code soon on github.com

Lovro Šubelj
University of Ljubljana
lovro.subelj@fri.uni-lj.si
http://lovro.lpt.fri.uni-lj.si

Ludo Waltman
Leiden University
waltmanlr@cwts.leidenuniv.nl
www.ludowaltman.nl

Vincent Traag
Leiden University
v.a.traag@cwts.leidenuniv.nl
www.traag.net

Nees Jan van Eck
Leiden University
ecknjpvan@cwts.leidenuniv.nl
www.neesjanvaneck.nl

CWTS
Meaningful metrics