intermediacy of publications

Lovro Šubelj
University of Ljubljana
Faculty of Computer and Information Science

Ludo Waltman
Leiden University
Centre for Science and Technology Studies

Vincent Traag
Leiden University
Centre for Science and Technology Studies

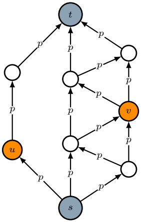
Nees Jan van Eck
Leiden University
Centre for Science and Technology Studies

NetSci '18

introduction & motivation

algorithmic historiography for evolution of field (**Garfield et al., 2003**)

relying on **citations** between scientific **publications** from **WoS** & **Scopus**



existing approaches include **main paths** (**Hummon & Doreian, 1989**)

(**longest/shortest paths**) many **irrelevant**/miss **relevant** publications

(**intermediacy**) important publications should only be **well-connected**

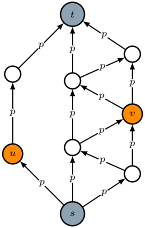
intermediacy measure

(**input**) selected **source** & **target** publications **s** & **t**

(**method**) each citation is relevant/active with **probability p**

(**measure**) importance of **publication u** called **intermediacy ϕ_u**

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

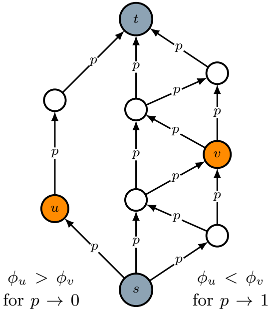


X_{st} exists path **from s to t** & **X_{st}^u** exists path **through u**

intermediacy for $p \to 0$

for $p \to 0$ intermediacy ϕ **governed by** ℓ (**proof**)

for $p \to 0$ if $\ell_u < \ell_v$ then $\phi_u > \phi_v$

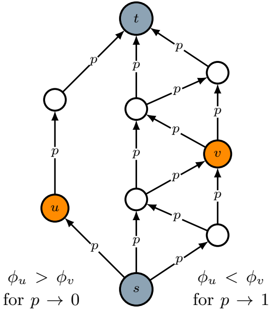


ℓ_u is **length** of **shortest paths** from s to t through u

intermediacy for $p \to 1$

for $p \to 1$ intermediacy ϕ **governed by σ** (**proof**)

for $p \to 1$ if $\sigma_u < \sigma_v$ then $\phi_u < \phi_v$

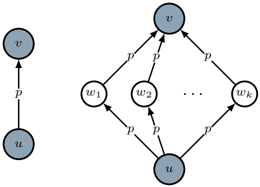


σ_u is **number** of **independent paths** from s to t through u

intuition for p

for what p is **direct citation** equivalent to **k indirect citations**

$$\Pr(X_{uv}) = p = 1 - (1 - p^2)^k$$



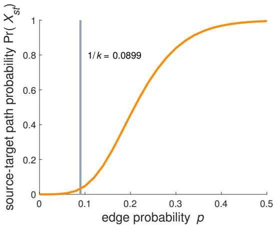
k	p
2	0.62
3	0.39
4	0.28
5	0.22
6	0.18
7	0.15
8	0.13
9	0.12
10	0.11

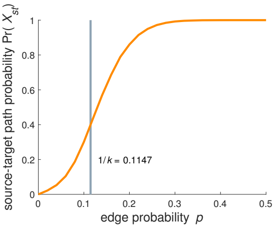
k is **number** of **independent paths** from u to v

phase transition

for what p source-target path $\mathbf{Pr(X_{st})} > \mathbf{0}$ & intermediacy $\exists\mathbf{u} : \phi_{\mathbf{u}} > \mathbf{0}$

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



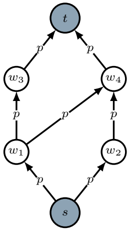


k is **average** number of **citations**/**references**

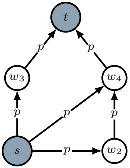
exact algorithm

decomposition algorithm by edge **contraction** & **removal** (**Ball, 1979**)

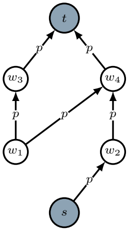
$$\Pr(X_{st} \mid G) = p \Pr(X_{st} \mid G/e) + (1-p) \Pr(X_{st} \mid G-e)$$



$$= \quad p$$



$$+ \quad (1-p)$$

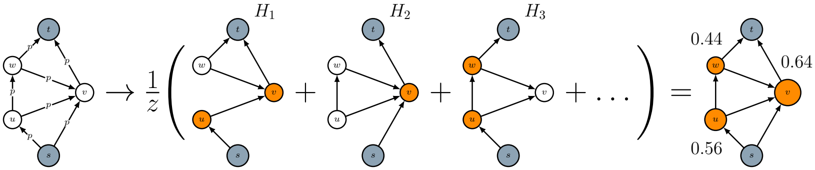


runs in **exponential time** since NP-hard even in DAG (**Johnson, 1984**)

approximate algorithm

simple **Monte Carlo** simulation algorithm by edge **sampling**

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



runs in quasi **linear time** using p-DFS over say **10^6 samples**

intermediacy $\neq$ centrality

correlation coefficient between **intermediacies ϕ** & **citations**/**references**

p = 0.1	1.00	0.69	0.28	0.16	0.12	0.19	0.20
p = 0.3	0.69	1.00	0.78	0.58	0.44	0.49	0.43
p = 0.5	0.28	0.78	1.00	0.92	0.77	0.50	0.36
p = 0.7	0.16	0.58	0.92	1.00	0.94	0.43	0.26
p = 0.9	0.12	0.44	0.77	0.94	1.00	0.35	0.19
cit.	0.19	0.49	0.50	0.43	0.35	1.00	0.00
ref.	0.20	0.43	0.36	0.26	0.19	0.00	1.00

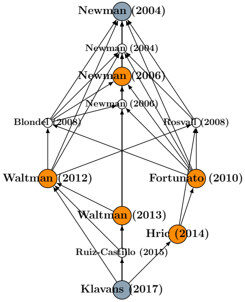
p = 0.1	1.00	0.42	0.21	0.10	0.04	0.16	0.03
p = 0.3	0.42	1.00	0.80	0.45	0.19	0.19	0.10
p = 0.5	0.21	0.80	1.00	0.78	0.39	0.18	0.14
p = 0.7	0.10	0.45	0.78	1.00	0.77	0.26	0.23
p = 0.9	0.04	0.19	0.39	0.77	1.00	0.26	0.24
cit.	0.16	0.19	0.18	0.26	0.26	1.00	0.01
ref.	0.03	0.10	0.14	0.23	0.24	0.01	1.00

intermediacy ϕ **uncorrelated** with standard **centrality measures**

modularity example

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

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



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.

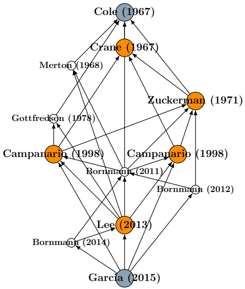
9 Newman (2004), Fast algorithm for detecting community structure in networks, *Phys. Rev. E* **69**(6), 066133.

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

peer review example

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

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



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.

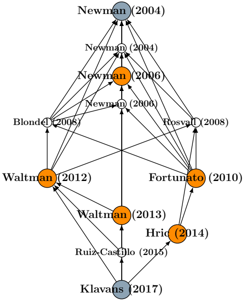
conclusions & future work

(**proposal**) measure of importance of publications called **intermediacy**

(**theory**) conceptually clear & provable behavior in **extreme cases**

(**practice**) intermediacy shows promising results in **case studies**

(**future**) applicability on general (**un**)**directed networks**?



(**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