intermediacy of **publications**

"pomembnost prispevkov za razvoj znanstvene tematike"

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

AI seminar '19

summer research visits


CWTS in Leiden, Vincent, Ludo & Nees

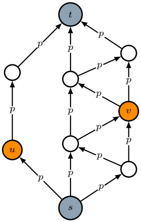



Tjaša, Nevi & Mažo, biking, boating & skating etc.

introduction & motivation

algorithmic historiography for evolution of field (**Garfield, 1964–**)

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**

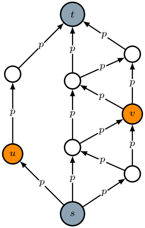
intermediacy measure

(**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

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

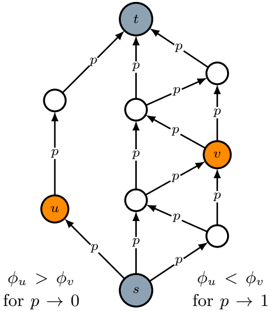


$\mathbf{X_{st}}$ exists path **from s to t** & $\mathbf{X_{st}^u}$ exists path **through u**

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 $\phi_u > \phi_v$

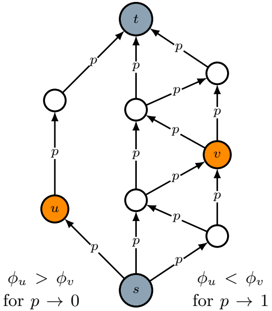


ℓ_u is **length** of **shortest paths** from s to t through u

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 $\phi_u < \phi_v$

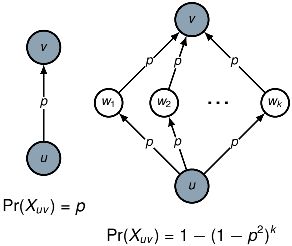


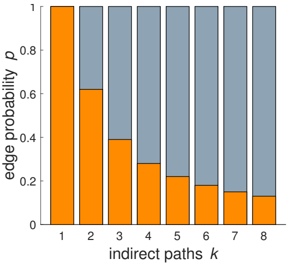
σ_u is **number** of **edge-disjoint paths** from s to t through u

intuition for p

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

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



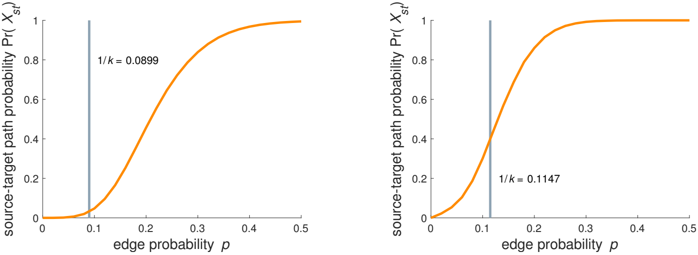


k is **number** of **indirect paths** from u to v

p phase transition

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

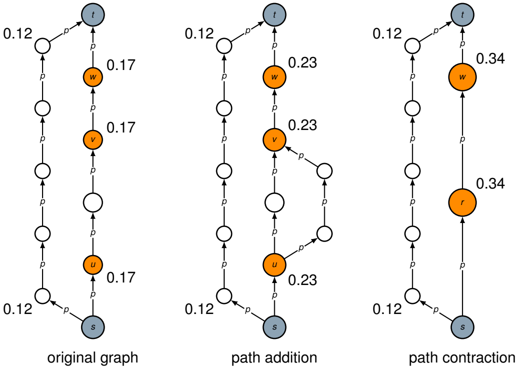
$$p \geq n/2m = 1/k$$



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

properties of intermediacy

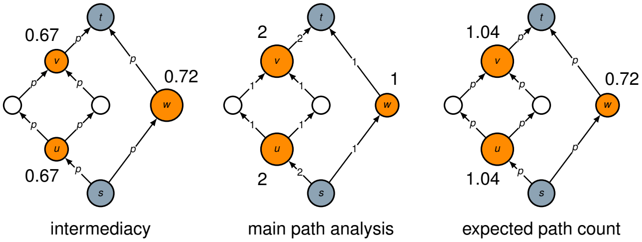
path **addition** & **contraction** increase intermediacy (**proof**)



path from source to target becomes **"easier"** (**intuition**)

alternatives to intermediacy

alternatives are **main paths** & **expected paths** (**state of the art**)

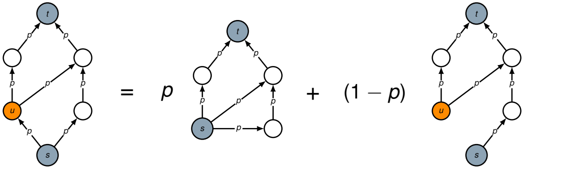


alternatives **violate** path contraction property (**example**)

exact algorithm

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

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



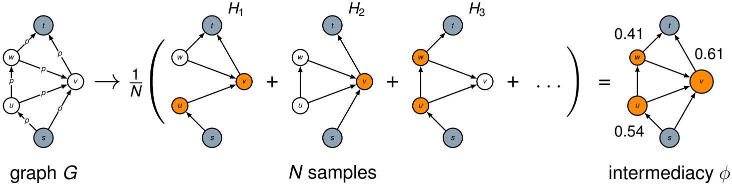
graph G contraction $G/(s,u)$ removal $G - (s,u)$

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

approximate algorithm

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

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



runs in quasi **linear time** using probabilistic 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

Pearson correlation

$p = 0.1$	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

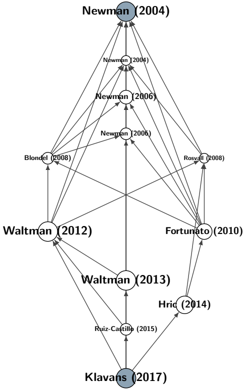
Pearson correlation

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.

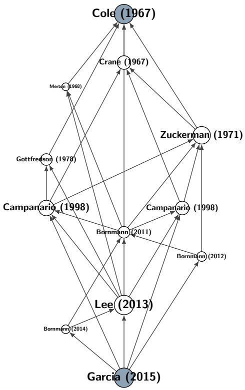
10 Rosvall & Bergstrom (2008), Maps of random walks on complex networks reveal community structure, *PNAS* **105**(4), 1118-1123.

in-house version of Scopus database at **CWTS**

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.

snapshot of WoS collected by (**Batagelj et al., 2017**)

sensing example

(**target**) Eagle & Pentland (2006), **Reality mining: Sensing complex social systems**, *Pers. Ubiquit. Comp.* **10**(4), 255-268.

(**source**) Mohr et al. (2017), **Personal sensing: Understanding mental health using ubiquitous sensors and machine learning**, *Annu. Rev. Clin. Psychol.* **13**, 23-47.

1 Eagle et al. (2009), Inferring friendship network structure by using mobile phone data, *PNAS* **106**(36), 15274-15278.

2 Miller (2012), The smartphone psychology manifesto, *Perspect. Psychol. Sci.* **7**(3), 221-237.

3 de Montjoye et al. (2013), Unique in the crowd: The privacy bounds of human mobility, *Sci. Rep.* **3**, 1376.

4 Andrews et al. (2015), Beyond self-report: Tools to compare estimated and real-world smartphone use, *PLoS ONE* **10**(10), e0139004.

5 Song et al. (2010), Limits of predictability in human mobility, *Science* **327**(5968), 1018-1021.

6 Gravenhorst et al. (2015), Mobile phones as medical devices in mental disorder treatment: An overview, *Pers. Ubiquit. Comp.* **19**(2), 335-353.

7 Grünerbl et al. (2015), Smartphone-based recognition of states and state changes in bipolar disorder patients, *IEEE J. Biomed. Health* **19**(1), 140-148.

8 Lane et al. (2010), A survey of mobile phone sensing, *IEEE Commun. Mag.* **48**(9), 140-150.

9 Eagle (2008), Behavioral inference across cultures: Using telephones as a cultural lens, *IEEE Intell. Syst.* **23**(4), 62-64.

10 Piwek et al. (2016), The rise of consumer health wearables: Promises and barriers, *PLoS Med.* **13**(2), e1001953.

suggested by **Veljko Pejović** for Scopus database

rendering example

(**target**) Wenger et al. (2004), **Interactive volume rendering of thin thread structures within multivalued scientific data sets**, *IEEE T. Vis. Comput. Gr.* **10**(6), 664-672.

(**source**) Eid et al. (2017), **Cinematic rendering in CT: A novel, lifelike 3D visualization technique,** *Am. J. Roentgenol.* **209**(2), 370-379.

1 Zhang et al. (2011), Volume visualization: A technical overview with a focus on medical applications, *J. Digit. Imaging* **24**(4), 640-664.

2 Bruckner & Gröller (2007), Enhancing depth-perception with flexible volumetric halos, *IEEE T. Vis. Comput. Gr.* **13**(6), 1344-1351.

3 Boucheny et al. (2009), A perceptive evaluation of volume rendering techniques, *ACM T. Appl. Percept.* **5**(4), 23.

4 Svakhine et al. (2009), Illustration-inspired depth enhanced volumetric medical visualization, *IEEE T. Vis. Comput. Gr.* **15**(1), 77-86.

5 Kroes et al. (2012), Exposure render: An interactive photo-realistic volume rendering framework, *PLoS ONE* **7**(7), e38586.

6 Dappa et al. (2016), Cinematic rendering: An alternative to volume rendering for 3D computed tomography imaging, *Insights Imaging* **7**(6), 849-856.

7 Schlegel et al. (2011), Extinction-based shading and illumination in GPU volume ray-casting, *IEEE T. Vis. Comput. Gr.* **17**(12), 1795-1802.

8 Schott et al. (2009), A directional occlusion shading model for interactive direct volume rendering, *Comput. Graph. Forum* **28**(3), 855-862.

9 Šoltészová et al. (2010), A multidirectional occlusion shading model for direct volume rendering, *Comput. Graph. Forum* **29**(3), 883-891.

10 Díaz et al. (2010), Real-time ambient occlusion and halos with Summed Area Tables, *Comput. Graph.* **34**(4), 337-350.

suggested by **Ciril Bohak** for Scopus database

chess example

(**target**) Guid & Bratko (2006), **Computer analysis of world chess champions**, *ICGA J.* **29**(2), 65-73.

(**source**) Mohr et al. (2017), **Personal sensing: Understanding mental health using ubiquitous sensors and machine learning**, *Annu. Rev. Clin. Psychol.* **13**, 23-47.

1 Guid & Bratko (2011), Using heuristic-search based engines for estimating human skill at chess, *ICGA J.* **34**(2), 71-81.

2 Guid et al. (2008), How trustworthy is CRAFTY's analysis of world chess champions?, *ICGA J.* **31**(3), 131-144.

3 Ferreira (2012), Determining the strength of chess players based on actual play, *ICGA J.* **35**(1), 3-19.

4 Barnes & Hernandez-Castro (2015), On the limits of engine analysis for cheating detection in chess, *Comput. Secur.* **48**, 58-73.

5 Guid & Bratko (2007), Factors affecting diminishing returns for searching deeper, *ICGA J.* **30**(2), 75-84.

6 Dailey et al. (2014), Move similarity analysis in chess programs, *Entertain. Comput.* **5**(3), 159-171.

7 Haworth (2007), Gentlemen, stop your engines!, *ICGA J.* **30**(3), 150-156.

8 Ferreira (2013), The impact of search depth on chess playing strength, *ICGA J.* **36**(2), 67-80.

suggested by **Matej Guid** for Scopus database

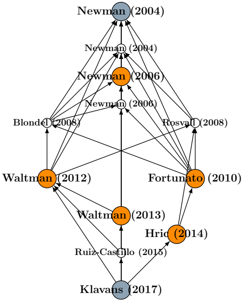
conclusions & future work

(**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**) online **research app**! applicability to **other networks**?



(**paper**) arxiv.org/abs/**1812.08259**

(**code**) github.com/lovre/**intermediacy**

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
http://www.ludowaltman.nl

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

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