intermediacy of **publications**

uncovering important publications for the development of a field

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

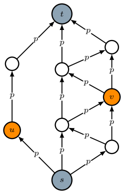
Nees Jan van Eck
Leiden University
Centre for Science and Technology Studies

CEU '25

problem & motivation

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

relying on **citations between publications** from **WoS/Scopus**



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

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

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

"...citations are valid and valuable means of creating accurate historical descriptions of scientific fields."

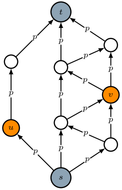
measure of intermediacy

(**setting**) select **source** & **target** publications **s** & **t**

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

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

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



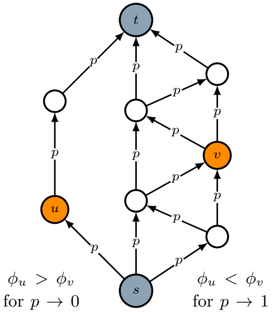
$\mathbf{X_{st}}$ – exists path **from s to t** & $\mathbf{X^u_{st}}$ – exists such path **through u**

$\phi_u = 2\phi_v \not\equiv$ publication u is "twice" as important as publication v

limit case $p \to 0$

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

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

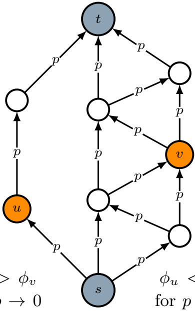


ℓ_u – **length** of **shortest paths** from s to t through u

limit case $p \to 1$

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

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



$\phi_u > \phi_v$ for $p \to 0$

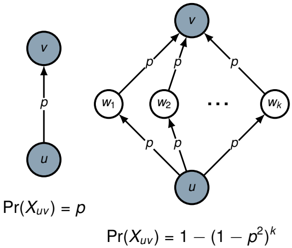
$\phi_u < \phi_v$ for $p \to 1$

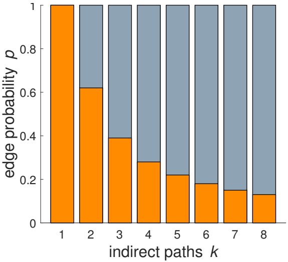
σ_u – **number** of **edge-disjoint paths** from s to t through u

intuition for parameter p

for what p is **direct citation** $\equiv$ k **indirect citations**

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





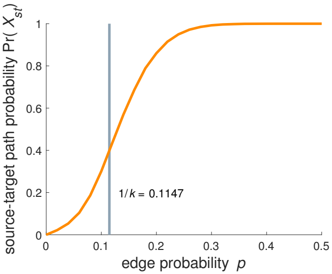
k – **number** of **indirect paths** from u to v

for example $p = 0.22 \equiv k = 5$ & $p = 0.11 \equiv k = 10$

choice of parameter p

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

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



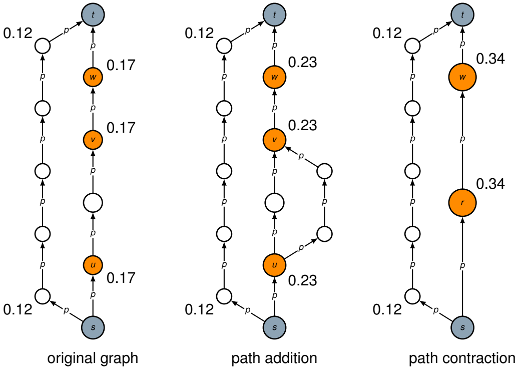
$\langle k \rangle$ – **average** number of **citations** & **references**

---

percolation theory suggests that for $k > 1$ probability $\Pr(X_{st})$ is non-negligible

properties of intermediacy

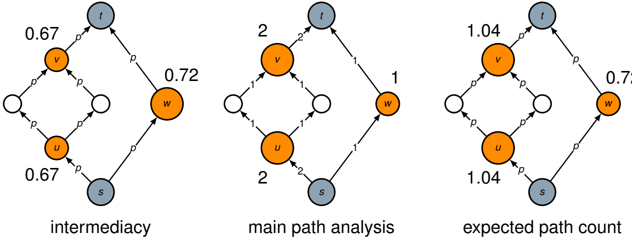
path addition & **contraction** increase intermediacy (**proof**)



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

alternatives to intermediacy

alternatives include **main paths** & **resistance** (**state of the art**)

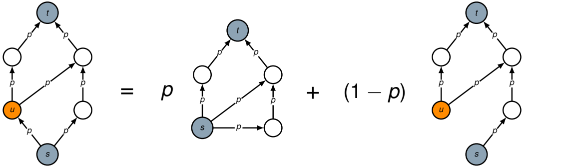


alternatives **violate** path addition/contraction property (**examples**)

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**)

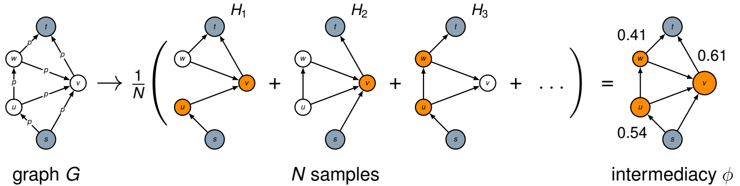
---

decomposition algorithm is cute & elegant, but not useful in practice :(

approximate algorithm

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

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

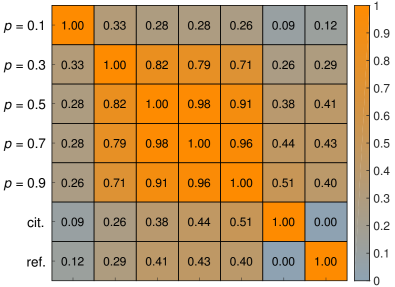


runs in quasi **linear time** using probabilistic DFS over say **10^6 samples**

$\ll$ 30 min for network with 9 145 771 nodes and 81 771 723 edges :)

intermediacy $\neq$ centrality

Spearman correlation between **intermediacy ϕ** & **citations/references**



intermediacy **not correlated** with standard **centrality measures**

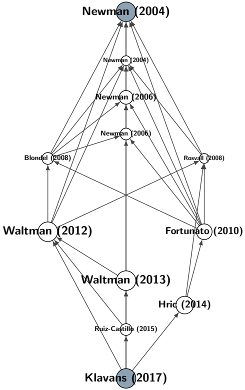
---

intermediacy most useful from ordinal perspective $\equiv$ Pearson $<$ Spearman correlation

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.

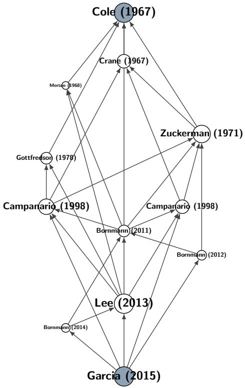
---

we set $p = 0.1$ & use 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.

we set $p = 0.1$ & use snapshot of **WoS** collected by (**Batagelj et al., 2017**)

small-world example

(**target**) Watts & Strogatz (1998), **Collective dynamics of 'small-world' networks**, *Nature* **393**(6684), 440-442.

(**source**) Backstrom et al. (2012), **Four degrees of separation**, In: *Proceedings of the WebSci '12*, pp. 45-54.

1 Newman (2003), The structure and function of complex networks, *SIAM Rev.* **45**(2), 167-256.

2 Albert & Barabási (2002), Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74**(1), 47-97.

3 Li et al. (2005), Towards a theory of scale-free graphs: Definition, properties, and implications, *Internet Math.* **2**(4), 431-523.

4 Leskovec et al. (2007), Graph evolution: Densification and shrinking diameters, *ACM Trans. Knowl. Discov. Data* **1**(1), 1-41.

5 Liben-Nowell et al. (2005), Geographic routing in social networks, *P. Natl. Acad. Sci. USA* **102**(33), 11623-11628.

6 Strogatz (2001), Exploring complex networks, *Nature* **410**(6825), 268-276.

7 Boldi et al. (2011), Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks, In: *Proceedings of the WWW '11*, pp. 587-596.

8 Dorogovtsev (2002), Evolution of networks, *Adv. Phys.* **51**(4), 1079-1187.

9 Ye et al. (2010), Distance distribution and average shortest path length estimation in real-world networks, In: *Proceedings of the ADMA '10*, pp. 322-333.

10 Lattanzi et al. (2011), Milgram-routing in social networks, In: *Proceedings of the WWW '11*, pp. 725-734.

---

we set $p = 0.1$ & use in-house version of **Scopus** database at **CWTS**

deep learning example

(**target**) LeCun et al. (2015), **Deep learning**, *Nature* **521**(7553), 436-444.

(**source**) Silver et al. (2017), **Mastering the game of Go without human knowledge**, *Nature* **550**(7676), 354-359.

1 Silver et al. (2016), Mastering the game of Go with deep neural networks and tree search, *Nature* **529**(7587), 484-489.
2 Jouppi et al. (2017), In-datacenter performance analysis of a tensor processing unit, In: *Proceedings of the ISCA '17*, pp. 1-12.
3 Reagen et al. (2016), Minerva: Enabling low-power, highly-accurate deep neural network accelerators, In: *Proceedings of the ISCA '16*, pp. 267-278.
4 Chen et al. (2016), DianNao family: Energy-efficient hardware accelerators for machine learning, *Commun. ACM* **59**(11), 105-112.
5 Chen et al. (2016), Eyeriss: A spatial architecture for energy-efficient dataflow for convolutional neural networks, In: *Proceedings of the ISCA '16*, pp. 127-138.
6 Moravčík et al. (2017), DeepStack: Expert-level artificial intelligence in heads-up no-limit poker, *Science* **356**(6337), 508-513.
7 Albericio et al. (2016), Cnvlutin: Ineffectual-neuron-free deep neural network computing, In: *Proceedings of the ISCA '16*, pp. 1-13.
8 Han et al. (2016), EIE: Efficient inference engine on compressed deep neural network, In: *Proceedings of the ISCA '16*, pp. 243-254.
9 Shafiee et al. (2016), ISAAC: A convolutional neural network accelerator with in-situ analog arithmetic in crossbars, In: *Proceedings of the ISCA '16*, pp. 14-26.
10 Adolf et al. (2016), Fathom: Reference workloads for modern deep learning methods, In: *Proceedings of the IISWC '16*, pp. 1-10.

---

we set $p = 0.1$ & use in-house version of Scopus database at **CWTS**

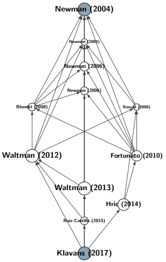
conclusions

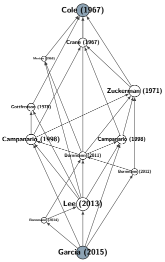
(**proposal**) measure of **importance of publications** called intermediacy

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

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

(**extensions**) multiple sources & targets, weighted networks, etc.





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

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

Šubelj, Waltman, Traag & Van Eck (2020) Intermediacy of publications, *Royal Society Open Science*, **7**(1), 190207.

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

convexity in complex networks

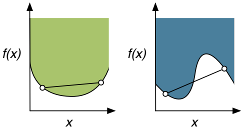
Tilen Marc
University of Ljubljana
Faculty of Mathematics and Physics

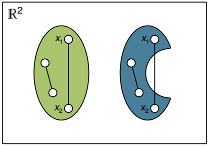
Lovro Šubelj
University of Ljubljana
Faculty of Computer and Information Science

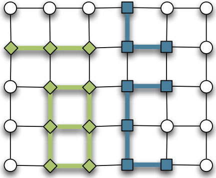
CEU '25

definitions of convexity

convex/**non-convex** real functions, sets in $\mathbb{R}^2$ & subgraphs

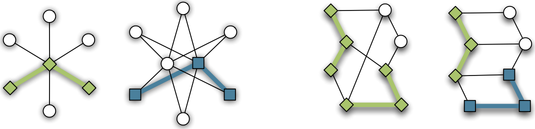






disconnected $\supseteq$ connected $\supseteq$ **induced** $\supseteq$ isometric $\supseteq$ **convex** subgraphs

connected subgraphs induced on simple undirected graph



convexity **in** networks?

(**social network analysis**) k-clubs/k-clans are convex k-cliques

(**community detection**) often defined as "convex" subgraph

— **subset** S is convex if it induces convex **subgraph**

— convex **hull** $\mathcal{H}(S)$ is smallest convex subset including S

hull number $= \min\{|S| : \mathcal{H}(S)$ includes n nodes$\}$ (**Everett & Seidman, 1985**)

↑ hull number measures how **quickly** convex subsets can grow

↓ how **slowly** randomly grown convex subsets expand

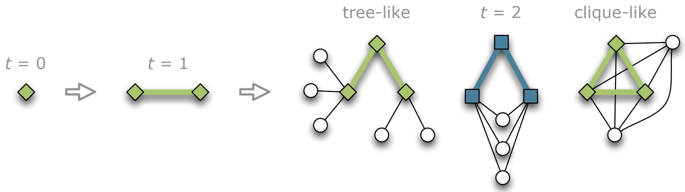
expansion of convex subsets

grow subset S by one node & **expand** S to convex hull $\mathcal{H}(S)$

— $S = \{\text{random node } i\}$

— until S contains n nodes:

1. select $i \notin S$ via random edge
2. expand $S = \mathcal{H}(S \cup \{i\})$

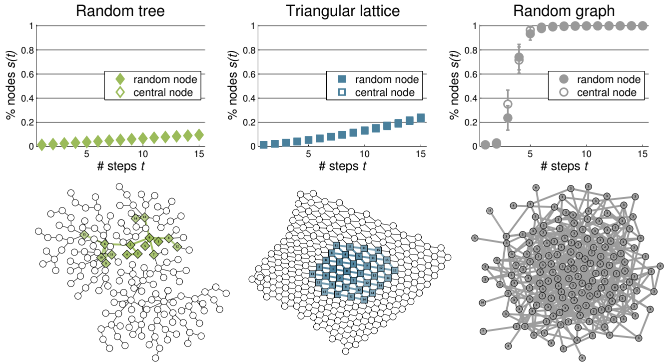


S quantifies (locally) **tree-like**/**clique-like** structure of graphs

convex expansion in graphs

$s(t)$ = average fraction of nodes in S after t expansion steps

$s(t) = (t+1)/n$ in **convex** & $s(t) \gg (t+1)/n$ in **non-convex** graphs

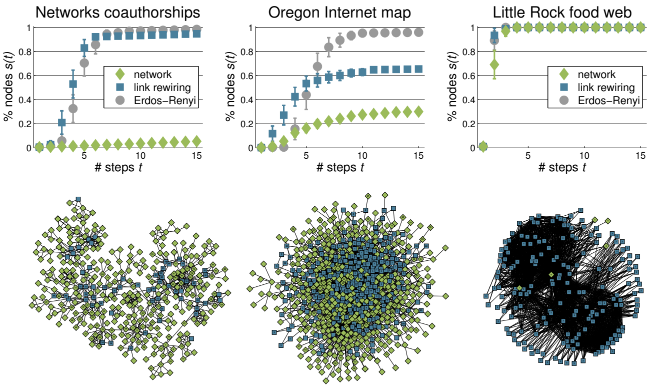


$s(t)$ quantifies (locally) **tree-like**/**clique-like** structure of graphs

convex expansion in networks

$s(t) =$ average fraction of nodes in S after t expansion steps

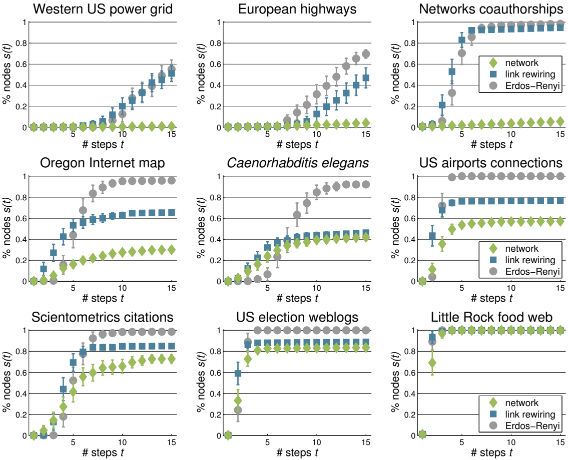
$s(t) = (t+1)/n$ in **convex** & $s(t) \gg (t+1)/n$ in **non-convex** networks



$s(t)$ quantifies (locally) **tree-like**/**clique-like** structure of networks

convex expansion in networks

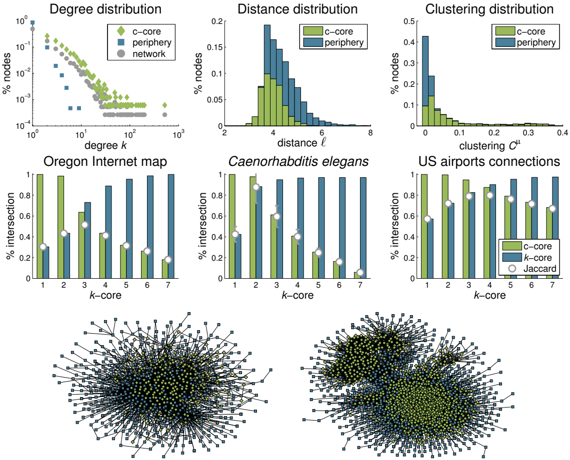
convex infrastructure and collaboration & **non-convex** food web



random **graphs** fail to reproduce convexity in empirical **networks**

when/why expansion settles?

(**when**) S extends to c-core (**why**) smallest convex subset $\supseteq S$



core-periphery networks have **convex** periphery & **non-convex** c-core

global measure c-convexity

$$X_c = 1 - \sum_{t=1}^{n-1} \sqrt[c]{\max(\Delta s(t) - 1/n, 0)} \qquad X_c \geq X_c^{\mathrm{RW}} \geq X_c^{\mathrm{ER}}$$

X_c highlights **tree-like**/**clique-like** networks (cliques connected tree-like)

	X_1	X_1^{RW}	X_1^{ER}	$X_{1.1}$	$X_{1.1}^{\mathrm{RW}}$	$X_{1.1}^{\mathrm{ER}}$
Western US power grid*	0.95	0.32	0.24	0.91	0.10	0.01
European highways*	0.66	0.23	0.27	0.44	−0.02	0.06
Networks coauthorships	0.91	0.09	0.06	0.83	−0.05	−0.09
Oregon Internet map	0.68	0.36	0.06	0.53	0.20	−0.09
Caenorhabditis elegans	0.57	0.54	0.07	0.43	0.40	−0.13
US airports connections	0.43	0.24	0.00	0.30	0.16	−0.07
Scientometrics citations	0.24	0.16	0.02	0.04	0.00	−0.13
US election weblogs	0.17	0.12	0.00	0.06	0.04	−0.08
Little Rock food web	0.03	0.03	0.02	−0.06	−0.02	−0.02

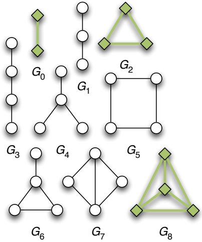
X_c measures **global** & **regional** (periphery) convexity in networks

probability of convex subgraphs

$P =$ probability that random $G_{1\text{-}8}$ convex $\qquad P \leq P^{\mathrm{ER}}$

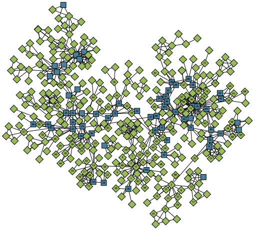
P highlights locally **tree-like**/**clique-like** networks & random graphs

	P	P^{ER}	$\ln n / \ln \langle k \rangle$
Western US power grid	77.0%	99.4%	8.66
European highways	83.2%	97.6%	7.54
Networks coauthorships	53.3%	71.3%	3.77
Oregon Internet map	56.0%	86.4%	4.40
Caenorhabditis elegans	77.8%	97.6%	5.79
US airports connections	5.5%	12.9%	2.38
Scientometrics citations	30.5%	89.2%	4.30
US election weblogs	2.7%	6.0%	2.15
Little Rock food web	2.2%	0.3%	1.59



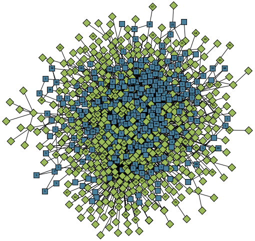
P measures **local** (up to **4 nodes**) convexity in networks

types of network convexity



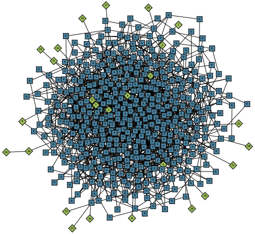
global convexity

tree/clique-like networks



regional convexity

core-periphery networks etc.



local convexity

random graphs

$< \ln n / \ln \langle k \rangle$

c-**convexity** $\neq$ standard measures & **c-core** $\neq$ k-cores

to be continued. . .

(**paper**) arxiv.org/abs/**1608.03402**

(**c++**) github.com/t4c1/**graph-convexity**

Marc & Šubelj (2018) Convexity in complex networks, *Network Science*, **6**(2), 176-203.

Tilen Marc
University of Ljubljana
tilen.marc@fmf.uni-lj.si
https://www.fmf.uni-lj.si

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

convex skeletons of networks

Lovro Šubelj
University of Ljubljana
Faculty of Computer and
Information Science

CEU ’25

corrected measure of c-convexity

$$Xs = s - \sum_{t=1}^{sn-1} \sqrt[c]{\max(s\Delta s(t) - 1/n, 0)} \qquad s = \text{fraction of nodes in LCC}$$

Xs highlights **tree-like**/**clique-like** networks & synthetic graphs

	n	$\langle k \rangle$	Xs
Jazz musicians	198	27.70	0.12
Network scientists	379	4.82	0.85
Computer scientists	239	4.75	0.64
Plasmodium falciparum	1158	4.15	0.43
Saccharomyces cerevisiae	1458	2.67	0.68
Caenorhabditis elegans	3747	4.14	0.56
AS (January 1, 1998)	3213	3.50	0.66
AS (January 1, 1999)	531	4.58	0.49
AS (January 1, 2000)	3570	3.94	0.59
Little Rock Lake	183	26.60	0.02
Florida Bay (wet)	128	32.42	0.03
Florida Bay (dry)	128	32.91	0.03

	n	$\langle k \rangle$	Xs
	2500	10.00	0.00
Random graphs	1000	10.00	0.01
	225	10.00	0.03
Triangular lattice	225	5.48	0.23
Rectangular lattice	225	3.73	0.13
Core-periphery graph	3747	4.48	0.39
	2500	5.97	1.00
Convex graphs	1000	5.97	1.00
	225	6.01	1.00

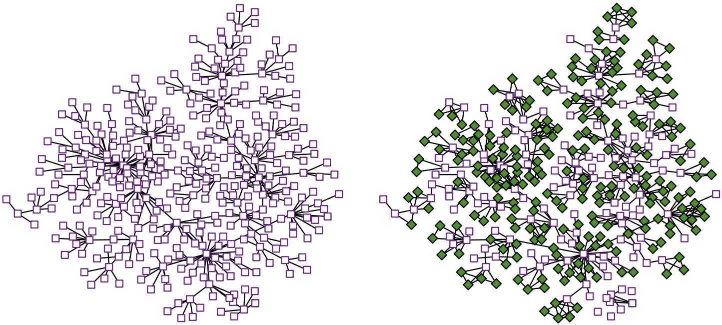
convex graphs are random trees of cliques

Xs measures **global** & **regional** convexity in (disconnected) networks

convex skeletons of networks

convex skeleton = largest high-*Xs* subnetwork (every *S* is convex)

spanning tree & **convex skeleton** of network scientists coauthorships



convex skeleton is **tree** of **cliques** extracted by edge removal

statistics of convex skeletons

$$\langle C \rangle = \frac{1}{n}\sum_i \frac{2t_i}{k_i(k_i-1)} \qquad \langle \sigma \rangle = \frac{2}{n(n-1)}\sum_{i<j}\sigma_{ij} \qquad Xs = \ldots$$

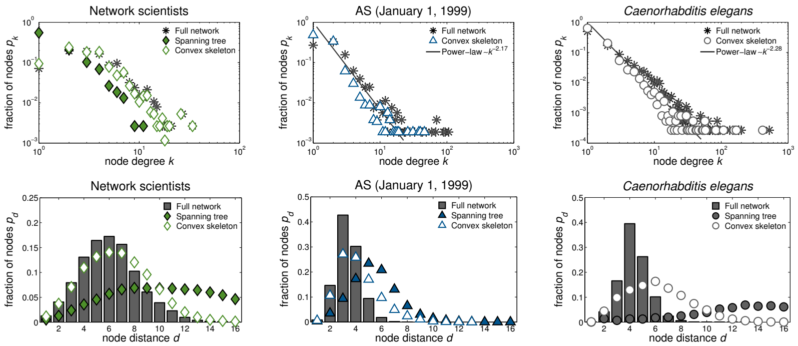
statistics of **convex skeletons** & **spanning trees** of networks

	clustering $\langle C \rangle$			geodesics $\langle \sigma \rangle$			convexity Xs		
	N	CS	ST	N	CS	ST	N	CS	ST
Jazz musicians	0.62	0.81	0.00	9.71	1.97	1.00	0.12	0.84	1.00
Network scientists	0.74	0.75	0.00	2.66	1.47	1.00	0.85	0.95	1.00
Computer scientists	0.48	0.54	0.00	4.08	1.42	1.00	0.64	0.95	1.00
Plasmodium falciparum	0.02	0.07	0.00	3.71	1.77	1.00	0.43	0.95	1.00
Saccharomyces cerevisiae	0.07	0.10	0.00	2.58	1.19	1.00	0.68	0.88	1.00
Caenorhabditis elegans	0.06	0.12	0.00	6.79	3.03	1.00	0.56	0.85	1.00
AS (January 1, 1998)	0.18	0.21	0.00	3.87	2.32	1.00	0.66	0.91	1.00
AS (January 1, 1999)	0.18	0.27	0.00	3.54	2.05	1.00	0.49	0.95	1.00
AS (January 1, 2000)	0.20	0.25	0.00	4.81	3.07	1.00	0.59	0.90	1.00
Little Rock Lake	0.32	0.69	0.00	22.13	4.32	1.00	0.02	0.82	1.00
Florida Bay (wet)	0.33	0.79	0.00	9.17	1.37	1.00	0.03	0.92	1.00
Florida Bay (dry)	0.33	0.82	0.00	9.37	1.65	1.00	0.03	0.93	1.00

convex skeleton is generalization of **spanning tree** retaining **clustering**

distributions of convex skeletons

node distributions of **convex skeletons** & **spanning trees** of networks

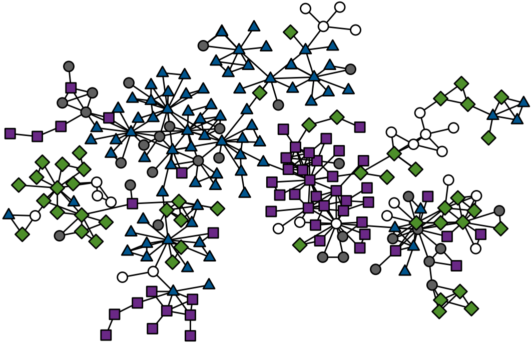


convex skeletons retain distributions in contrast to **spanning trees**

convex skeletons of coauthorships

convex skeleton ∼ network abstraction technique

convex skeleton of Slovenian **computer scientists** coauthorships



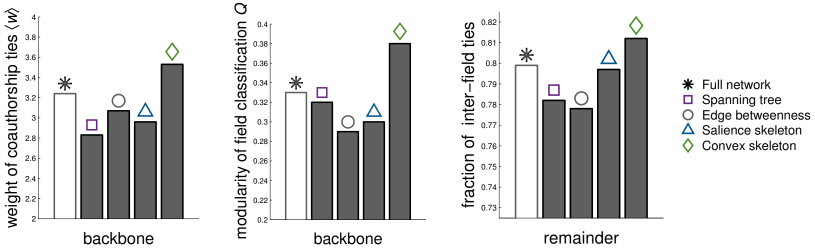
computer theory (◆), information systems (■), intelligent systems (▲),

programming technologies (○) & other (●)

network backbones of coauthorships

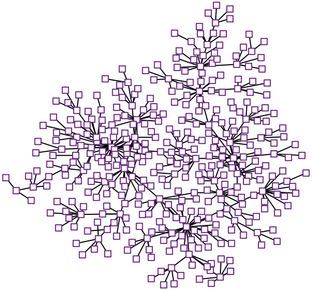
convex skeleton $\gg$ high-betweenness & high-salience skeletons

properties of **backbones** of Slovenian **computer scientists** coauthorships



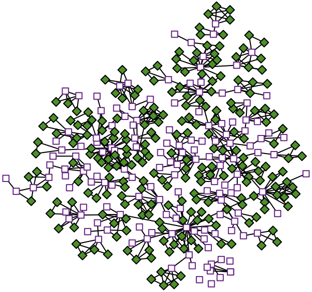
convex skeletons increase properties in contrast to **other backbones**

convex skeletons of networks



spanning **tree**

tree w/o cliques



convex skeleton

tree w/ cliques

convex skeleton $\gg$ backbones & **c-centrality** $\neq$ centralities

thank you!

(**paper**) arxiv.org/abs/**1709.00255**

(**c++**) github.com/t4c1/**convex-skeleton**

Šubelj (2018) Convex skeletons of complex networks, *Journal of the Royal Society Interface*, **15**(145), 20180422.

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