convexity in complex networks

Lovro Šubelj
University of Ljubljana
Faculty of Computer and Information Science

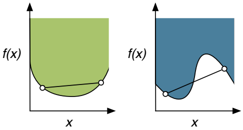
joint work with

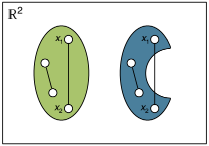
Tilen Marc
University of Ljubljana
Institute of Mathematics, Physics and Mechanics

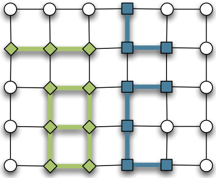
LCN2 '17

definitions of convexity

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

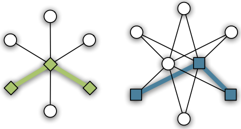


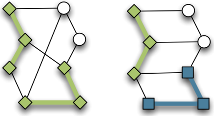




disconnected ⊇ connected ⊇ **induced** ⊇ isometric ⊇ **convex** subgraphs

connected subgraphs induced on simple undirected graph →





convexity **in** networks?

(**sna**) k-clubs/k-clans are convex k-cliques

(**cd**) community 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

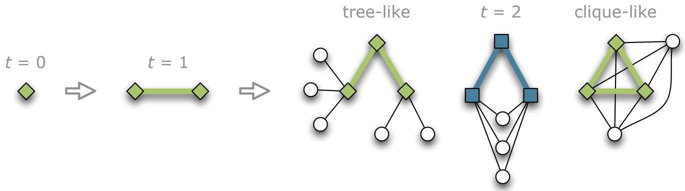
expansion of convex subsets

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

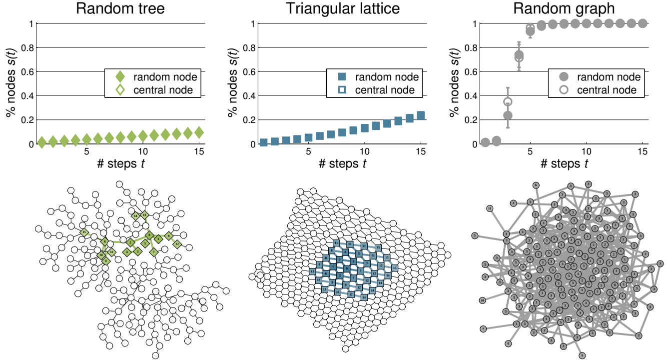


S quantifies (locally) **tree-like**/**clique-like** structure of graphs

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

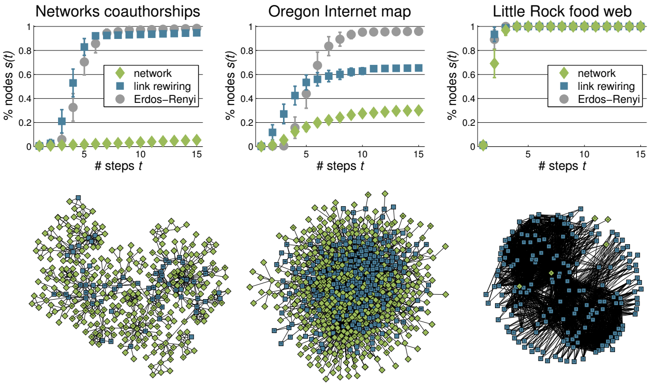


$s(t)$ quantifies (locally) **tree-like**/**clique-like** structure of graphs

convex expansion in networks

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

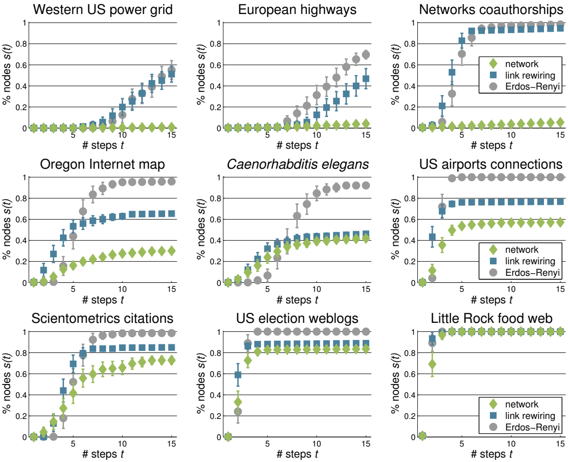
$s(t) = (t+1)/n$ in **convex** & $s(t) \gg (t+1)/n$ in **non-convex** networks



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

convex expansion in networks

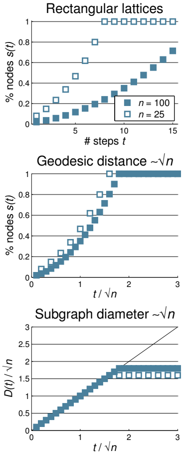
convex infrastructure and collaboration & non-convex food web



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

when/why sudden expansion?

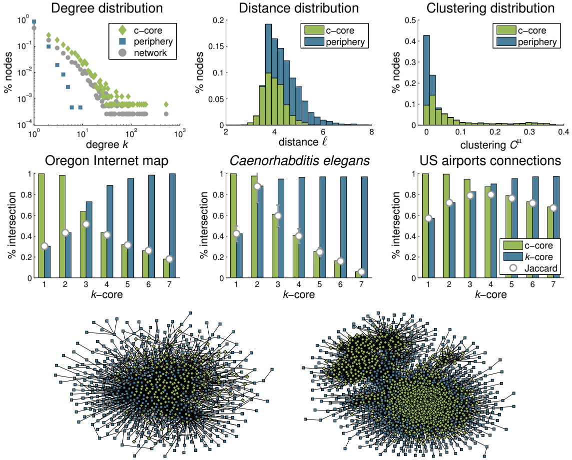
(**why**) steps $t \approx$ diameter $D(t) >$ distance $\langle \ell \rangle$ (**when**)



random graphs **convex** for $< \mathcal{O}(\ln n)$ & **non-convex** for $> \mathcal{O}(\ln^2 n)$

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

local measure of convexity

$$L_c = 1 + \max\{\, t \mid s(t) < (t+c+1)/n \,\} \qquad L_1 \leq L_1^{\mathrm{ER}} \approx \ln n / \ln \langle k \rangle$$

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

	L_t	L_t^{ER}	L_1	L_1^{ER}	$\ln n / \ln \langle k \rangle$
Western US power grid	14	9	6	9	8.66
European highways	16	7	7	7	7.54
Networks coauthorships	17	4	7	4	3.77
Oregon Internet map	3	4	3	4	4.40
Caenorhabditis elegans	2	5	2	5	5.79
US airports connections	2	3	2	3	2.38
Scientometrics citations	3	4	3	4	4.30
US election weblogs	2	2	2	2	2.15
Little Rock food web	2	2	2	2	1.59

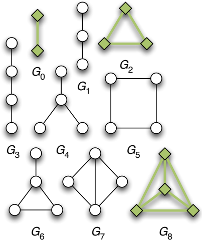
L_c measures **local** & **absolute** (tree/clique) convexity in networks

probability of convex subgraphs

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

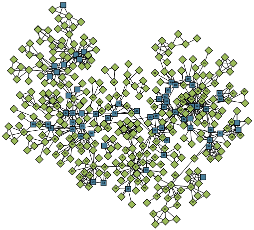
P highlights locally **tree-like**/**clique-like** networks & random graphs

	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



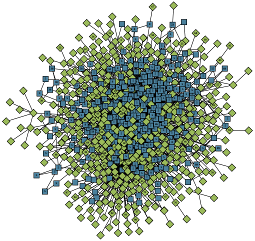
P measures **local** (up to **4 nodes**) convexity in networks

types of network convexity



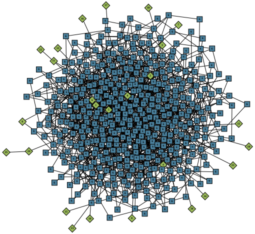
global convexity

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

robustness, navigation, optimization, abstraction, comparison etc.

to be continued...

arXiv:**1608.03402v3**

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

Tilen Marc
University of Ljubljana
tilen.marc@imfm.si
http://www.imfm.si

convex skeletons of networks

Lovro Šubelj
University of Ljubljana
Faculty of Computer and
Information Science

LCN2 '17

corrected measure of 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		n	$\langle k \rangle$	Xs
Jazz musicians	198	27.70	0.12		2500	10.00	0.00
Network scientists	379	4.82	0.85	Random graphs	1000	10.00	0.01
Computer scientists	239	4.75	0.64		225	10.00	0.03
Plasmodium falciparum	1158	4.15	0.43	Triangular lattice	225	5.48	0.23
Saccharomyces cerevisiae	1458	2.67	0.68	Rectangular lattice	225	3.73	0.13
Caenorhabditis elegans	3747	4.14	0.56	Core-periphery graph	3747	4.48	0.39
AS (January 1, 1998)	3213	3.50	0.66		2500	5.97	1.00
AS (January 1, 1999)	531	4.58	0.49	Convex graphs	1000	5.97	1.00
AS (January 1, 2000)	3570	3.94	0.59		225	6.01	1.00
Little Rock Lake	183	26.60	0.02				
Florida Bay (wet)	128	32.42	0.03				
Florida Bay (dry)	128	32.91	0.03				

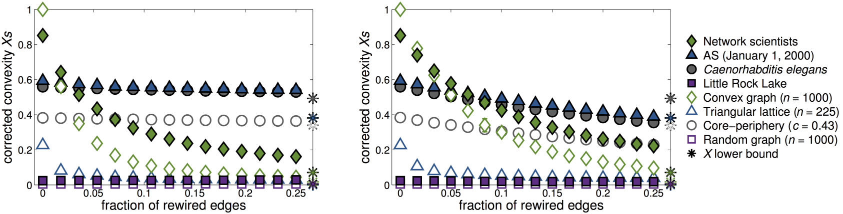
convex graphs are random trees of cliques

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

convexity under randomization

$X \geq s_1 \qquad s_1 =$ fraction of pendant nodes

Xs under **degree-preserving**/**full randomization** by edge rewiring

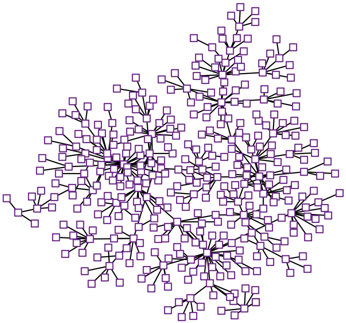


Xs very **sensitive** to **random perturbations** of network structure

convex skeletons of networks

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

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

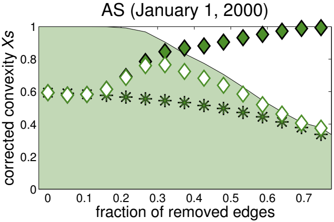


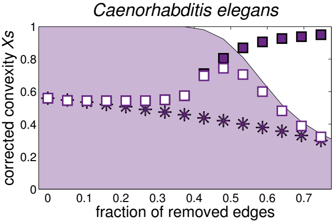
convex skeleton is **tree** of **cliques** extracted by edge removal

extraction of convex skeletons

$$c_i = \sum_{j \in \Gamma_i} p_j - \sum_{j \in \Gamma_i} 1 - p_j \qquad p_i = \text{probability that } i \in \text{c-core}$$

Xs under **removal** of edges $\{i, j\}$ based on **c-centrality** $c_i + c_j$





c-centrality $c_i + c_j$ for **core-periphery** & clustering $\Delta C_i + \Delta C_j$ for **others**

statistics of convex skeletons

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

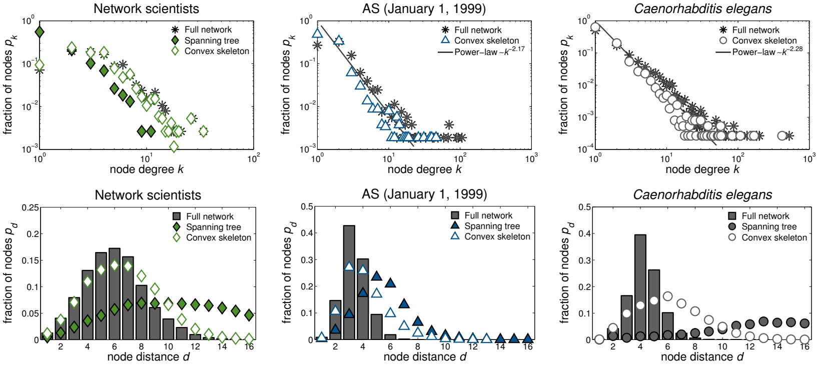
statistics of **convex skeletons** & **spanning trees** of networks

	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

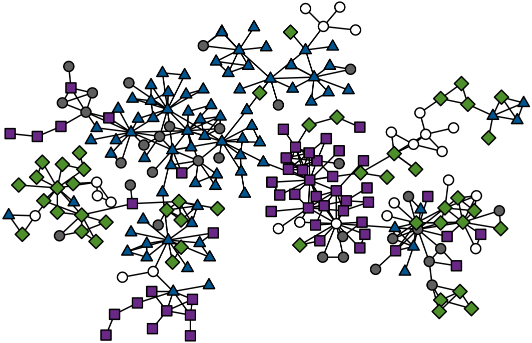


convex skeletons retain distributions in contrast to **spanning trees**

convex skeletons of coauthorships

convex skeleton ∼ network abstraction technique

convex skeleton of Slovenian **computer scientists** coauthorships

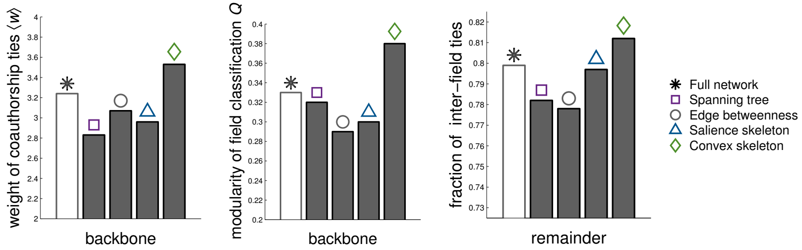


computer theory (◆), information systems (■), intelligent systems (▲), programming technologies (○) & other (●)

network backbones of coauthorships

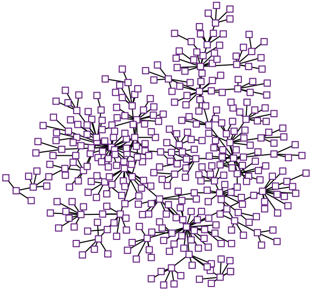
convex skeleton $\gg$ high-betweenness & high-salience skeletons

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



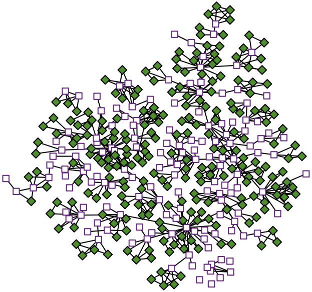
convex skeletons increase properties in contrast to **other backbones**

convex skeletons of networks



spanning **tree**

tree w/o cliques



convex skeleton

tree w/ cliques

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

abstraction, sampling, sparsification, modeling, dynamics etc.

thank you!

arXiv:**1709.00255v2**

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