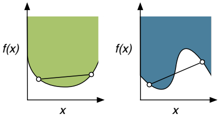
on convexity in complex networks

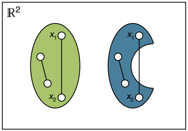
Tilen Marc (IMFM) & Lovro Šubelj (FRI)

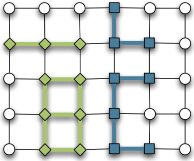
Wednesday seminar

definitions of convexity

- convex function, convex set and convex subgraph
- disconnected ⊇ connected ⊇ induced ⊇ isometric ⊇ convex
- we consider connected subgraphs induced on simple graph
- convex hull of subset of nodes and hull number (later on)

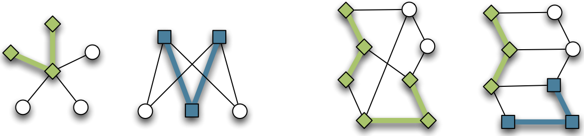






why interested in convexity?

— (left) hub-and-spokes arrangement and bipartite graph
— (right) two graphs identical up to 3-node subgraphs
— (why) possible applications in many network problems



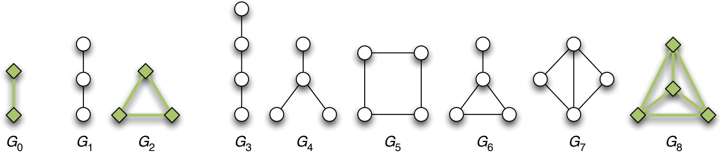
what is convexity in networks?

— k-cliques and k-clubs/clans in social network analysis

— definitions of groups of nodes in community detection

— frequency of motifs and graphlets in empirical networks

— (bottom) connected subgraphs with up to 4 nodes



convex hull and hull number

- let S be subset of nodes and $\mathcal{H}(S)$ its convex hull
- $\mathcal{H}(S)$ is smallest convex subgraph including S (unique)
- hull number is size of smallest S thus $\mathcal{H}(S)$ is entire graph

(draw picture)

- let S be convex subset if it induces convex subgraph
- hull number measures how quickly convex subsets can grow
- we study how slowly randomly grown convex subsets expand

expansion of convex subsets

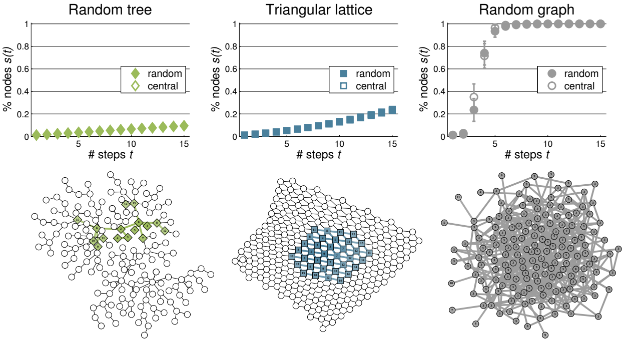
- grow subset S by one node
- expand S to convex hull $\mathcal{H}(S)$
- observe evolution of its size $|S|$

(draw picture)

1. select random seed i and set $S = \{i\}$
2. until S contains all nodes repeat:
   2.1 select $i \notin S$ by following random edge
   2.2 expand S to convex hull $\mathcal{H}(S \cup \{i\})$

expansion in graphs

- let $s(t)$ be fraction of nodes in S after t steps (2. step)
- $s(t)$ quantifies (locally) tree-like or clique-like structure
- let graph be convex if any S is convex, $s(t) = (t+1)/n$



networks and graphs

— nine empirical networks from various domains

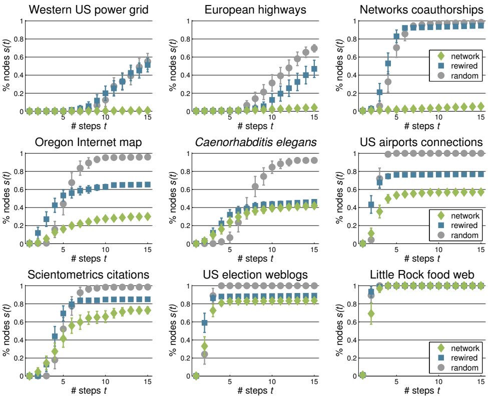
— Erdős-Rényi random graphs with same n and m

— configuration model graphs with same $k_1, k_2, \ldots, k_n$

network	nodes n	edges m	deg. $\langle k \rangle$	clus. $\langle C \rangle$	dist. $\langle \ell \rangle$
Western US power grid	4941	6594	2.67	0.08	18.99
European highways	1039	1305	2.51	0.02	18.40
Networks coauthorships	379	914	4.82	0.74	6.04
Scientometrics citations	1878	5412	5.76	0.13	5.52
Caenorhabditis elegans	3747	7762	4.14	0.06	4.32
US airports connections	1572	17214	21.90	0.50	3.12
Oregon Internet map	767	1734	4.52	0.29	3.03
US election weblogs	1222	16714	27.36	0.32	2.74
Little Rock food web	183	2434	26.60	0.32	2.15

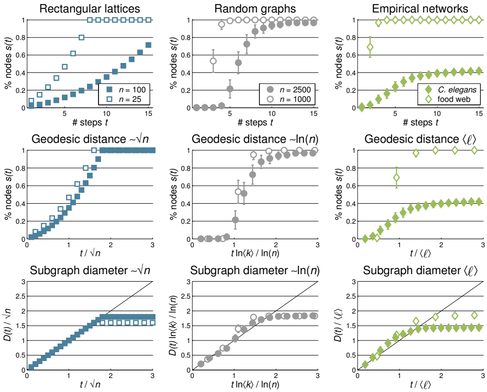
expansion in networks

- (convex) tree-like technological and clique-like collaboration
- (non-convex) food web, web graph and dense protein network
- (random) graphs fail to reproduce trends in empirical networks



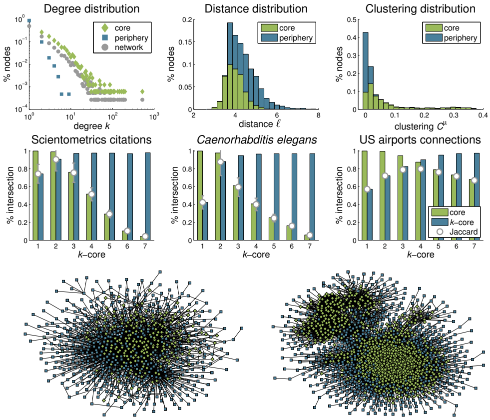
when/why sudden growth?

- (when) number of steps $t >$ average distance $\langle \ell \rangle$
- (why) subgraph diameter $D(t) >$ average distance $\langle \ell \rangle$
- (random) subset S is convex for $s < \log n$ when $n \to \infty$



when/why growth settles?

- (when) subset S extends to network core
- (why) core is smallest convex subset including S
- (networks) non-convex core and convex periphery



measure of convex growth

- let $\Delta s(t)$ be growth in tth step, $\Delta s(t) = s(t) - s(t-1)$
- let $\Delta\widetilde{s}(t)$ be growth in convex graph, $\Delta\widetilde{s}(t) = 1/n$

$$X_c = 1 - \sum_{t=1}^{n-1} \sqrt[c]{\max(\Delta s(t) - \Delta\widetilde{s}(t), 0)}$$

$$X_1 = \frac{\#\text{ steps to cover network} + 1}{n} \approx 1 - \frac{\#\text{ nodes in network core}}{n}$$

- X_c is approximated from 100 terms of sum and $c \geq 1$

convex growth in networks

— X_c highlights tree-like and/or clique-like networks
— X_c mixes local (random) and regional (periphery) convexity
— X_c wrongly estimates regional convexity (core-periphery)

network	X_1	$X_1^{\{k\}}$	X_1^m	$X_{1.1}$	$X_{1.1}^{\{k\}}$	$X_{1.1}^m$
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

local analysis of convexity

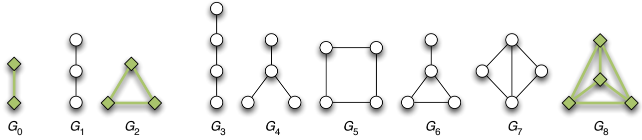
— (local) probability that induced subgraph is convex?

— (bottom) connected subgraphs G_i with up to 4 nodes

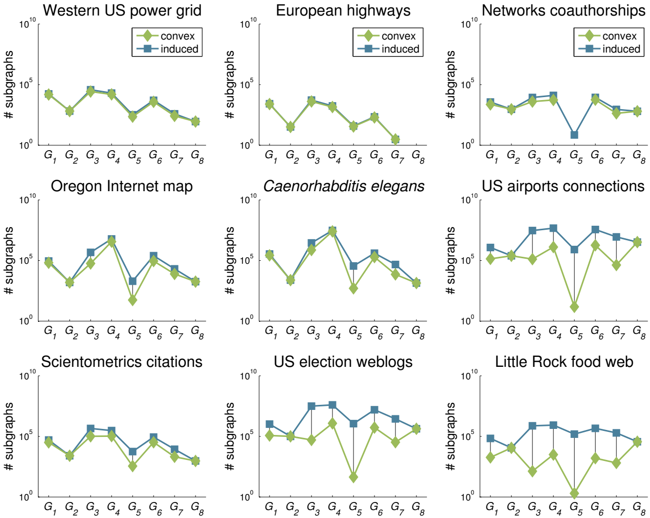
— (random) let P_i be probability that G_i is convex and p density

$$P_0 = 1 \quad P_1 = (1-p^2)^{n-3} \quad P_2 = 1 \quad \ldots$$

$$\ldots \quad P_6 = (1-2p^2+3p^3)^{n-4} \quad P_7 = (1-p^2)^{n-4} \quad P_8 = 1$$



frequency of convex subgraphs



probability of convex subgraph

- let g_i (c_i) be number of (convex) subgraphs G_i
- let P_c be probability that random G is convex
- probability P_c mostly consistent with measure X_c

$$P_c = \sum_i \frac{g_i}{\sum_i g_i} \frac{c_i}{g_i} = \frac{\sum_i c_i}{\sum_i g_i}$$

network	P_c
Western US power grid	77.0%
European highways	83.2%
Networks coauthorships	53.3%
Oregon Internet map	56.0%
Caenorhabditis elegans	77.8%
US airports connections	5.5%
Scientometrics citations	30.5%
US election weblogs	2.7%
Little Rock food web	2.2%

convexity in networks

— (convex) spatial technological $\approx$ social collaboration

— (non-convex) food web and weblogs graph

— (locally convex) only random graphs

— (application) sampling, comparison, navigation, redundancy

— (open) practical measure, non-simple networks etc.

arXiv:1608.03402