spanning trees that preserve network distances

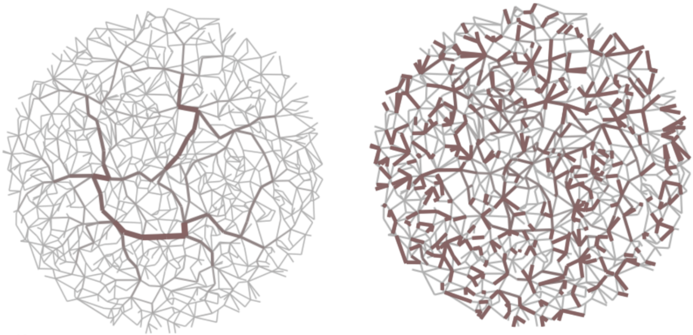
Lovro Šubelj
University of Ljubljana
Faculty of Computer and
Information Science

Networks '21

motivation

network abstraction with backbones and skeletons

(left) high-betweenness backbone and (right) high-salience skeleton (Grady et al., 2012)



spanning trees

network abstraction with **spanning trees**

consider connected unweighted network on n nodes

spanning tree is **connected with n nodes** and $n-1$ edges

trees lack clustering $\langle C \rangle = 0$ in contrast to convex skeletons (Šubelj, 2018)

are spanning trees also **small-world and scale-free?**

$\langle d \rangle \sim \log n$ in small-world networks and $p_k \sim k^{-\gamma}$ in scale-free networks

in random trees almost surely $\langle d \rangle \sim \sqrt{n}$ (Reńyi and Szekeres, 1967)

algorithms

Kruskal's algorithm

1. start with forest of trees each consisting of single node
2. merge trees until only one remains (using minimum edges)

Prim's algorithm

1. start with single tree consisting of (random) seed node
2. add one new node at each step (using minimum edge)

breadth-first search

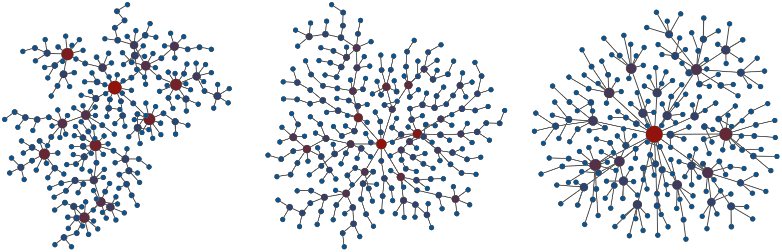
1. start with single tree consisting of (random) seed node
2. add new neighbors of node at each step (in breadth-first order)

other algorithms

depth-first search, Sollin's algorithm etc.

examples of spanning trees of random graph

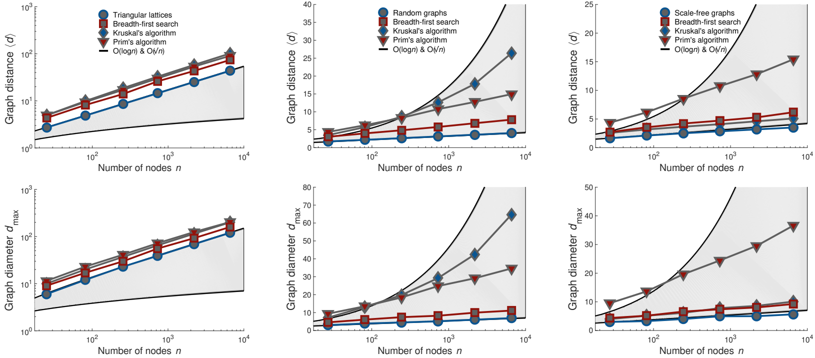
(left) Kruskal's algorithm, (middle) Prim's algorithm and (right) breadth-first search



distance $\langle d \rangle$ and diameter d_{max}

only **BFS retains** scaling of $\langle d \rangle$ **and** d_{max} in synthetic graphs

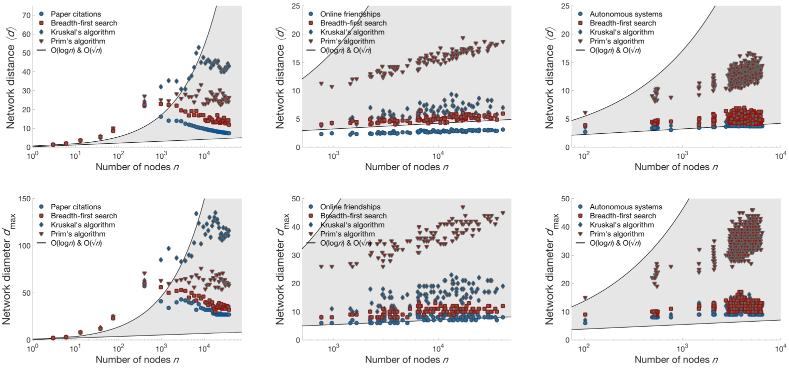
$\langle d \rangle \sim \sqrt{n}$ in lattices, $\langle d \rangle \sim \log n$ in random graphs and $\langle d \rangle \sim \frac{\log n}{\log \log n}$ in scale-free graphs



distance $\langle d \rangle$ and diameter d_{max}

only **BFS retains** scaling of $\langle d \rangle$ **and** d_{max} in real networks

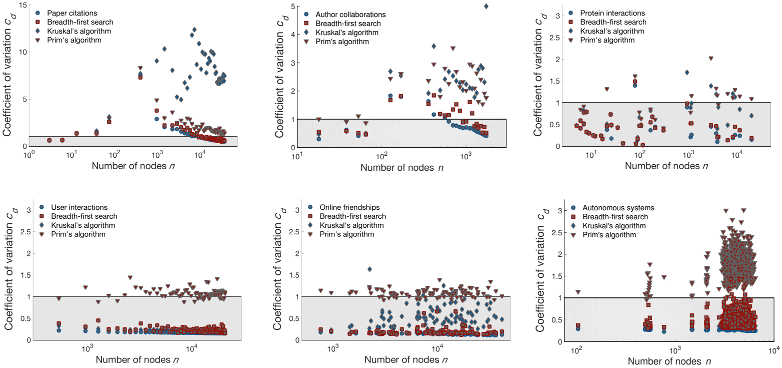
$\langle d \rangle \sim \log n$ in small-world networks and $\langle d \rangle \sim \log \log n$ in ultra small-world networks



distance distribution p_d

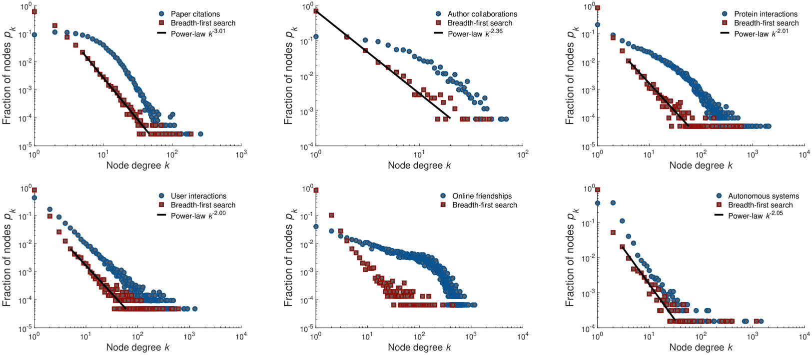
only **BFS retains low-variance p_d** in real networks

coefficient of variation $c_d = \frac{\sigma_d}{\langle d \rangle} < 1$ in (ultra) small-world networks



degree distribution p_k

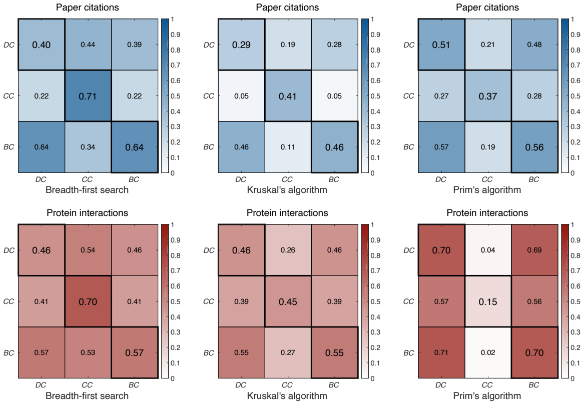
only BFS power-law $p_k \sim k^{-\gamma}$ in most cases (Clauset et al., 2009)



node importance

BFS best retains closeness centrality in real networks

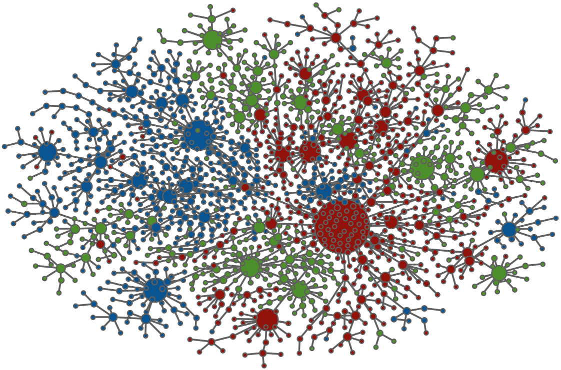
correlations between node degree (*DC*), closeness centrality (*CC*) and betweenness centrality (*BC*)



network visualization

BFS spanning tree of **author collaborations in Slovenia**

natural sciences (red), engineering (green), medical sciences (blue) and other



conclusions

spanning trees **small-world** $\langle d \rangle \sim \log n$ and **scale-free** $p_k \sim k^{-\gamma}$

trees lack clustering $\langle C \rangle = 0$ in contrast to convex skeletons (Šubelj, 2018)

use breadth-first search for unweighted networks!

use Prim's or Kruskal's algorithm only for weighted networks

are spanning trees actually **balanced trees?**

balanced tree data structure ensures $\langle d \rangle \approx \log n$ by definition

how to measure that tree is approximately balanced?

thank you!

Šubelj (2021) Algorithms for spanning trees of unweighted networks. *PeerJ Comput. Sci.*, under review.

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