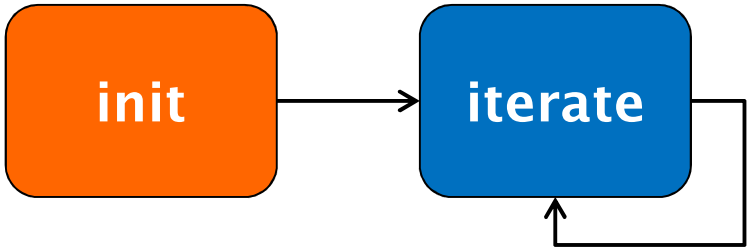

# Measuring Importance of Seeding for Structural De-anonymization Attacks in Social Networks

**Gábor György Gulyás*,** and Sándor Imre****
*Laboratory of Cryptography and System Security (CrySyS)
**Mobile Communications and Quantum Technologies Laboratory
Budapest University of Technology and Economics
gulyas@crysys.hu

# Outline and contributions

Powerful structural re-identification attacks

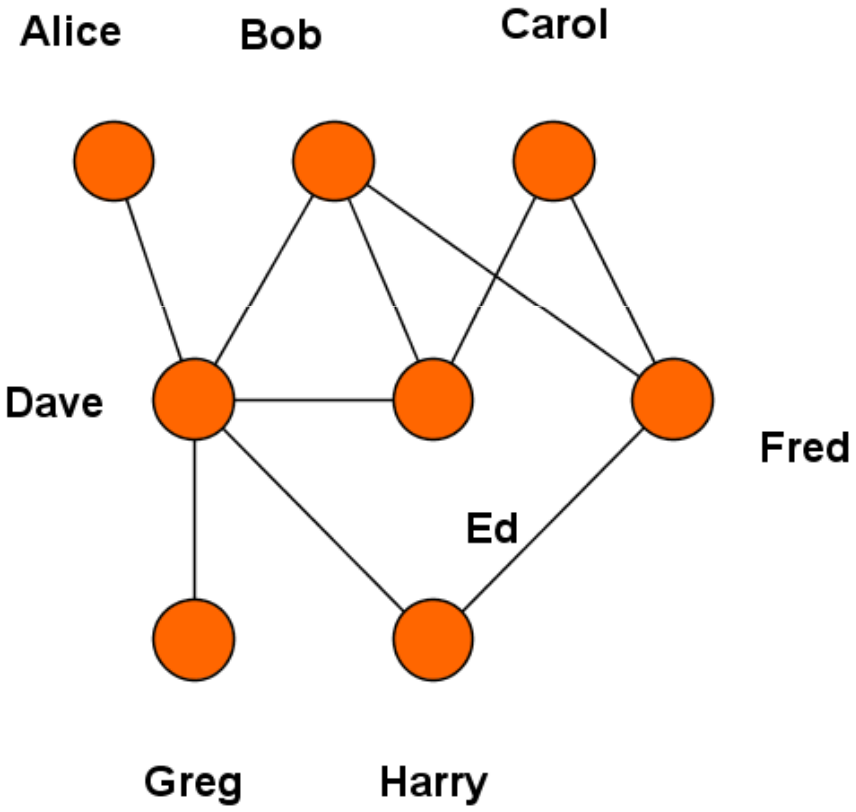



can achieve large-scale de-anonymization

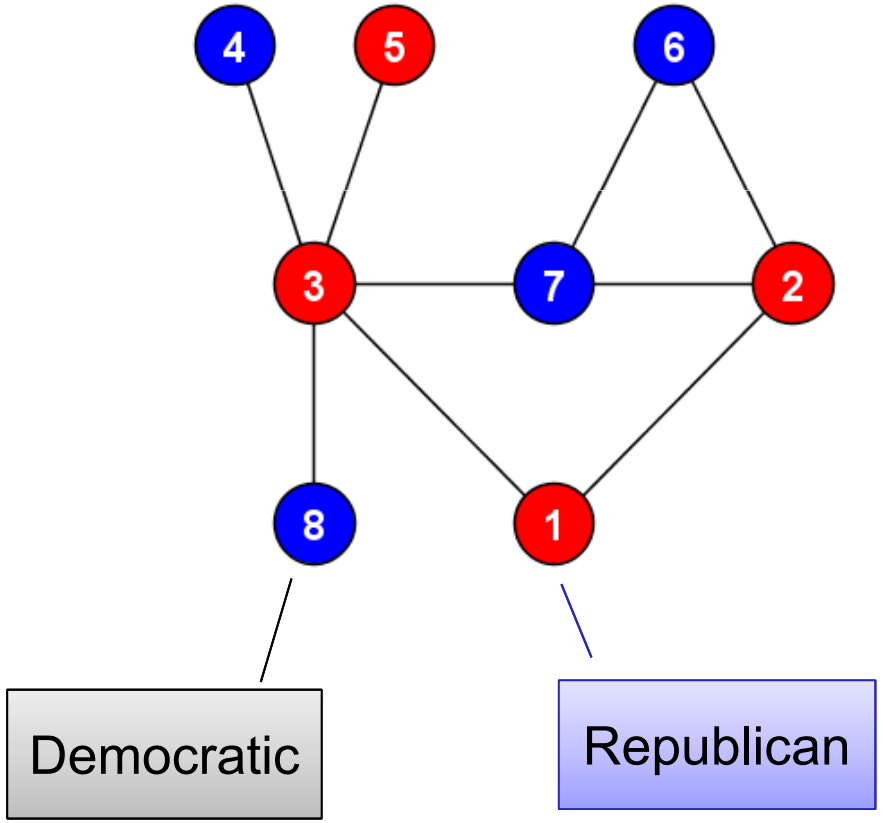
We analyze the init phase:
- compare several methods
- apsects wrt. network size, structure
- importance of seed node roles
- other interesting properties

# Re-identification in social networks

**Auxiliary information, $G_{src}$**
(a public crawl, e.g., Flickr)



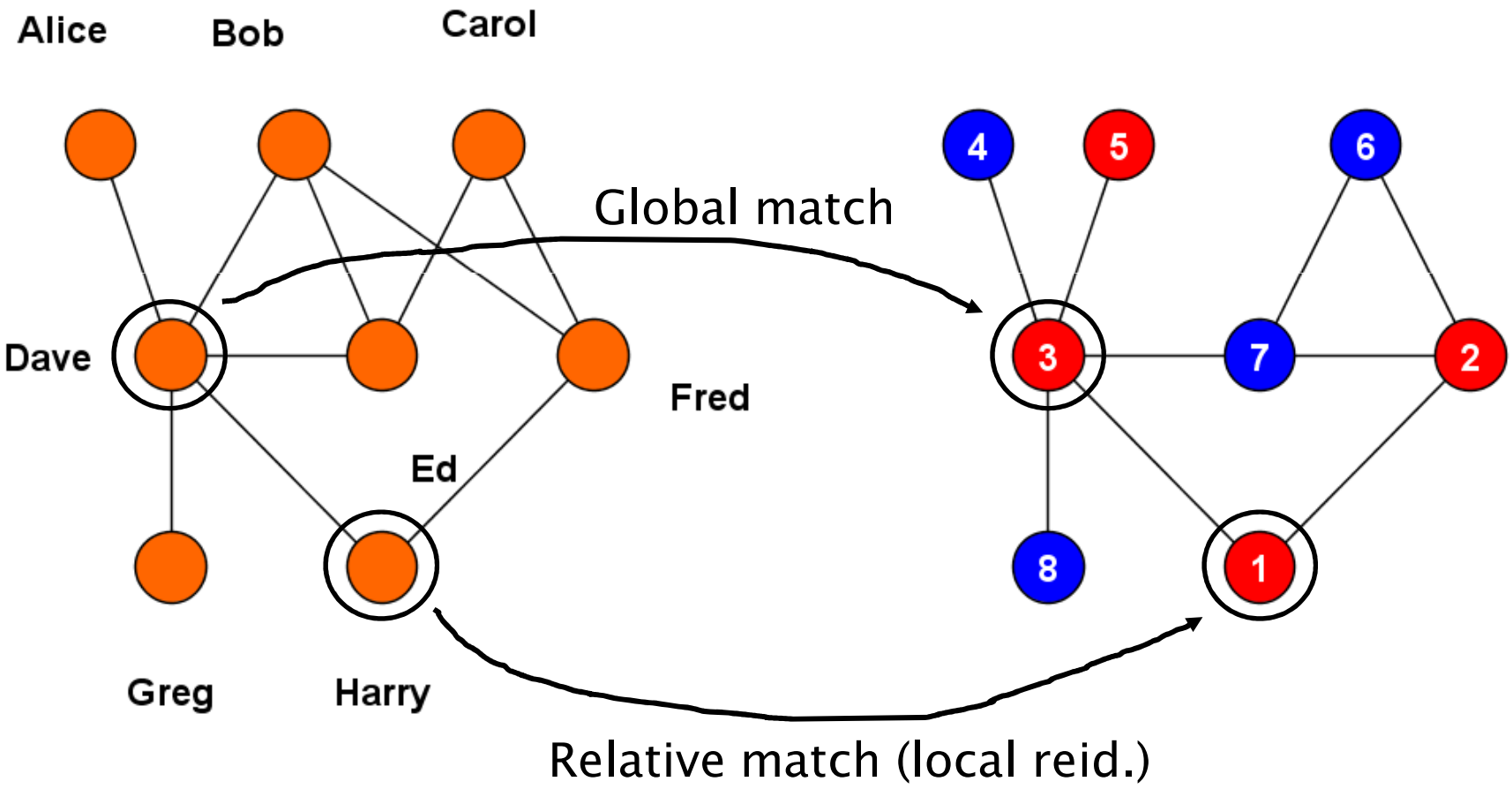

**Anonimized graph, $G_{tar}$**
(anonimized export, e.g., Twitter)



# Re-identification in social networks (2)

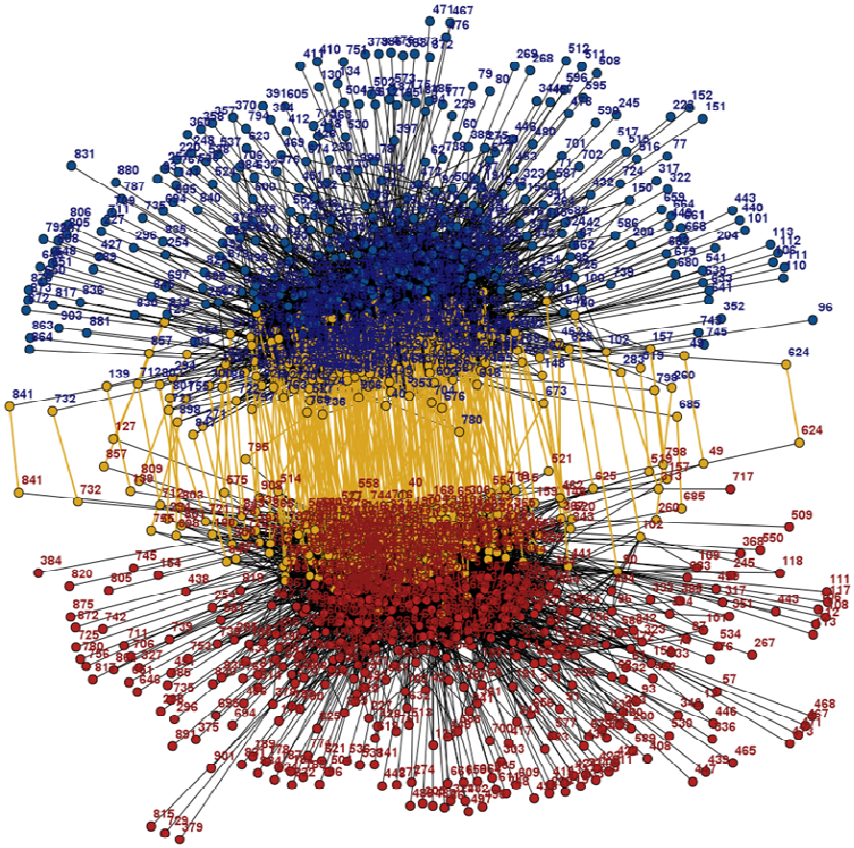
**Auxiliary information, $G_{src}$**
(a public crawl, e.g., Flickr)

**Anonimized graph, $G_{tar}$**
(anonimized export, e.g., Twitter)



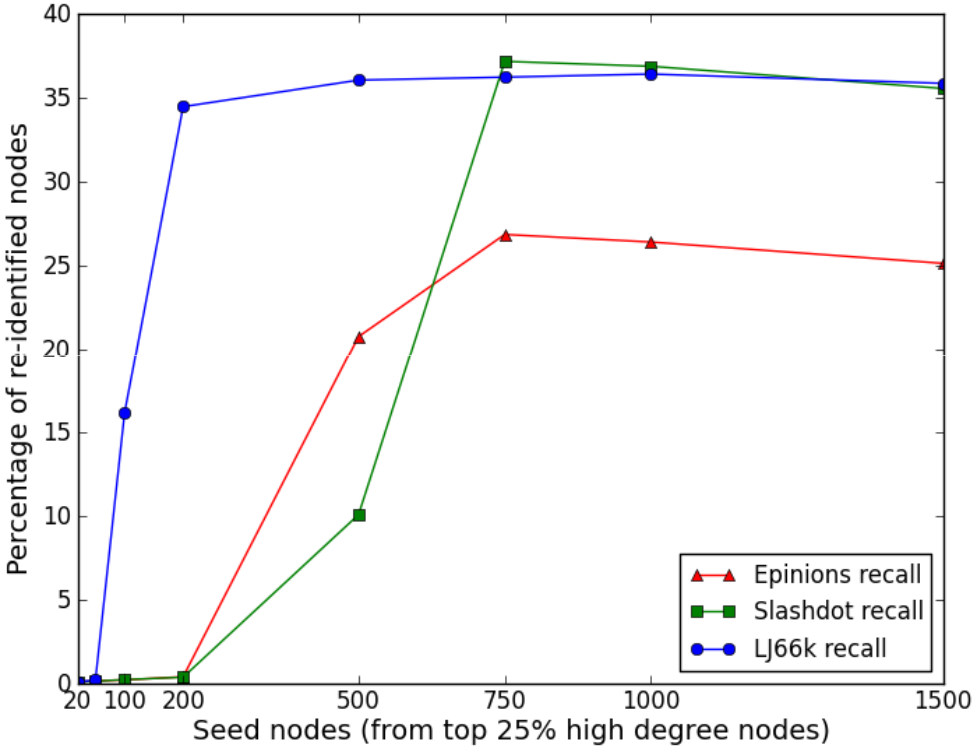
# Re-identification in social networks (3)

1. **Seed identification**
   - Kind of „initalization”
   - Global identification
2. **Propagation**
   - Starts from seed set
   - Local identification



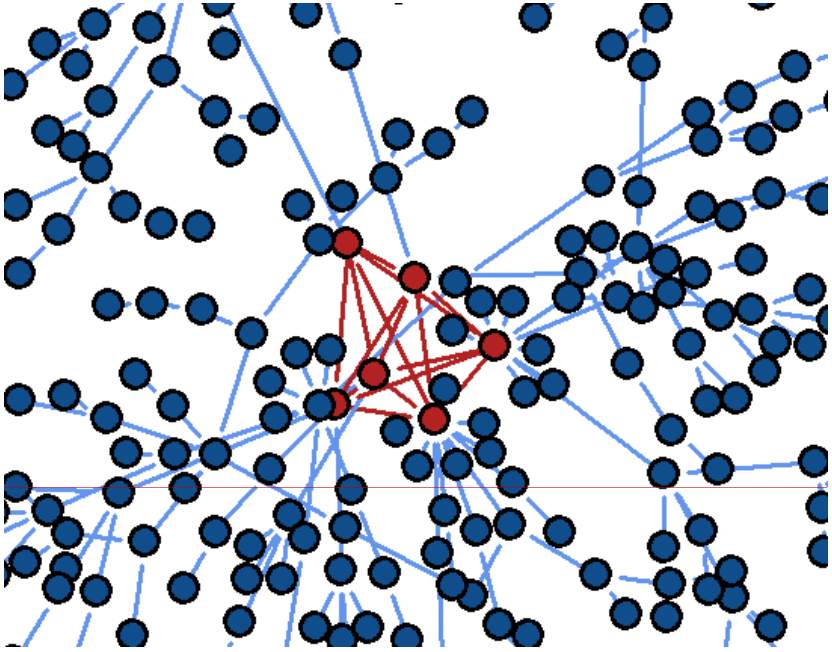
# Interesting properties of seeding in the literature

- Narayanan & Shmatikov, 2009
- Phase transition property, boundaries depend on:
  - Network structure
  - Seeding method
- Probability of propagation
  - Probability of wider propagation




(our measurements)

# Seeding in the literature



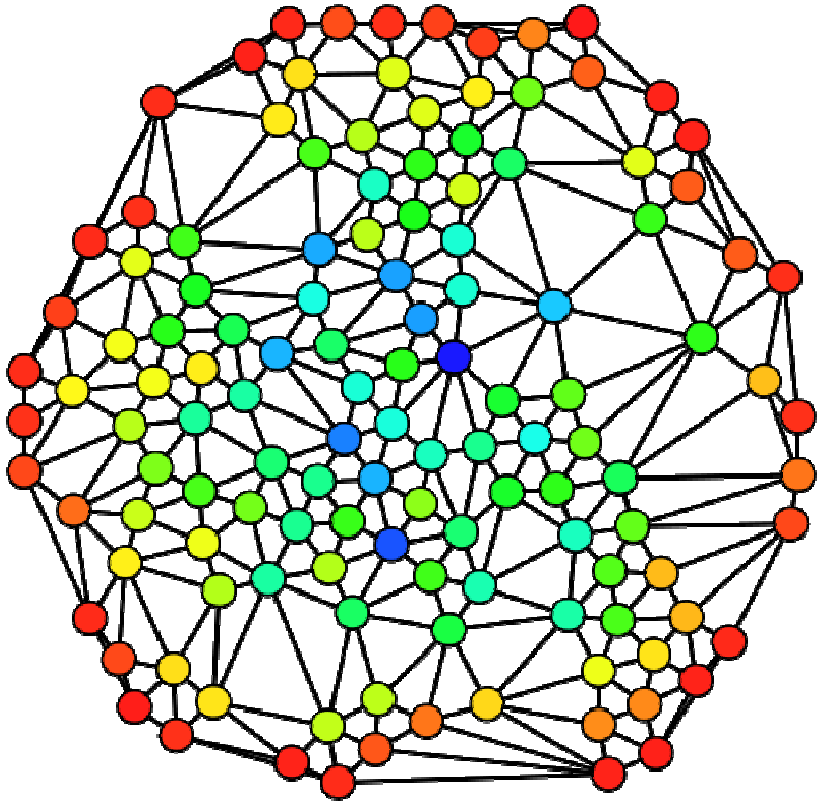
- Narayanan & Shmatikov, 2009:
  - Nodes of 4-cliques
  - High degree (min. 80)
- Narayanan et al., 2011
  - Top nodes by degree

***spoiler alert***

Our measurements verify that, interestingly, these are quite good choices.
However, this is not emphasized in the papers.

# Seeding in the literature (2)

- Srivatsa & Hicks, 2012
  - De-anonymizing mobility traces by using social networks
  - Betweenness centrality in social network
  - Avg. distance by probabilistics paths in the contact graph



Betweenness centrality (source: wikipedia)

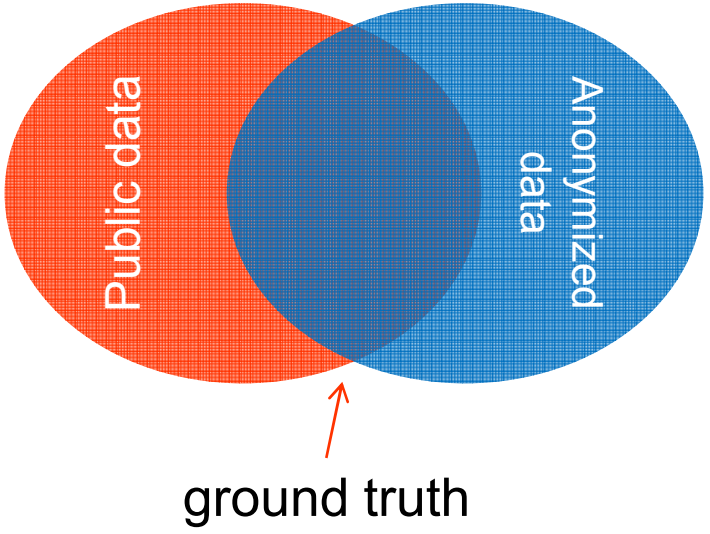
# Seeding in the literature (3)

- Wei et al., 2012
  - Seed and grow, an attack algorithm
  - Seed selection method is not detailed in the paper (i.e., how nodes selected that are identified by injected subgraph)
- Yartseva & Grossglauser, 2013
  - Formal analysis of a simplified attack variant (e.g., proving phase transition)
  - Seeding method is not detailed in the paper
- Beato et al., 2013 (protection)
  - Some nodes act as proxies to achieve structure modification
  - Top nodes selected as seeds
- Gulyas & Imre, 2013 (protection)
  - Privacy-enhancing identity management
  - Seeds are selected randomly from top 25% nodes (by deg.)

# Evaluation method

- Evaluation against the Nar09 attack (Narayanan & Shmatikov, 2009)
- Datasets: Slashdot, Epinions, LiveJournal
  - Structurally diverge
  - Large datasets (66-82k nodes), one smaller (10k nodes)
- Data perturbation
  - $\alpha_V$=0.5,
  - $\alpha_E$=0.75
- Attack parameters
  - 2 perturbation, 3 rounds
  - Θ = 0.01
  - Observed error rate ca. 1-2%

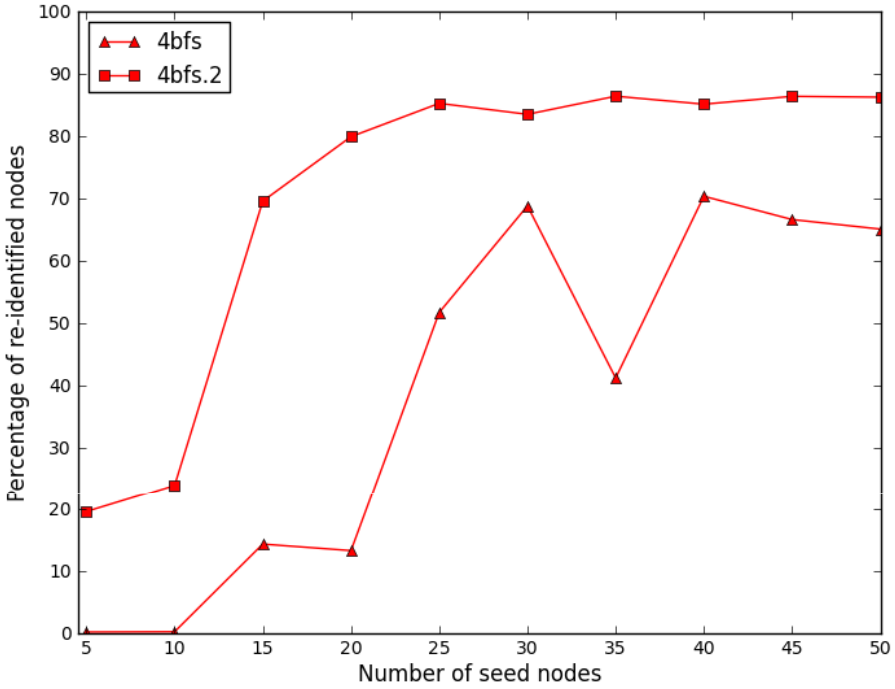
Java framework
ca. 6700 SLOC

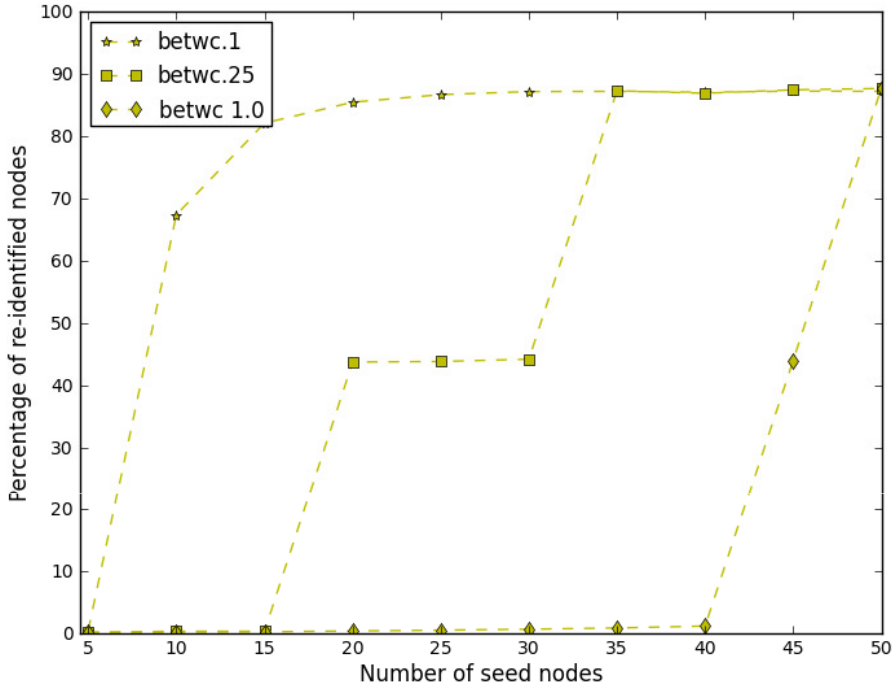


# Analyzed measures

- k-cliques ($k \in \{4, 5, 6\}$)
  - With any degree
  - From 20% top by deg.
- k-neighborhood
  - Similarily as above
- Random nodes
  - From 10%, 20%, 50% top by deg.
  - From all nodes
- Top nodes
- Betweenness centrality
  - With any degree
  - 10%, 20% top by deg
- Closeness centrality
  - Similarily as above
- Local Topological Anonymity (LTA)
- Local Clustering Coefficient (LCC)
  - LCCH: skipping top 20%

# Measurements of some properties

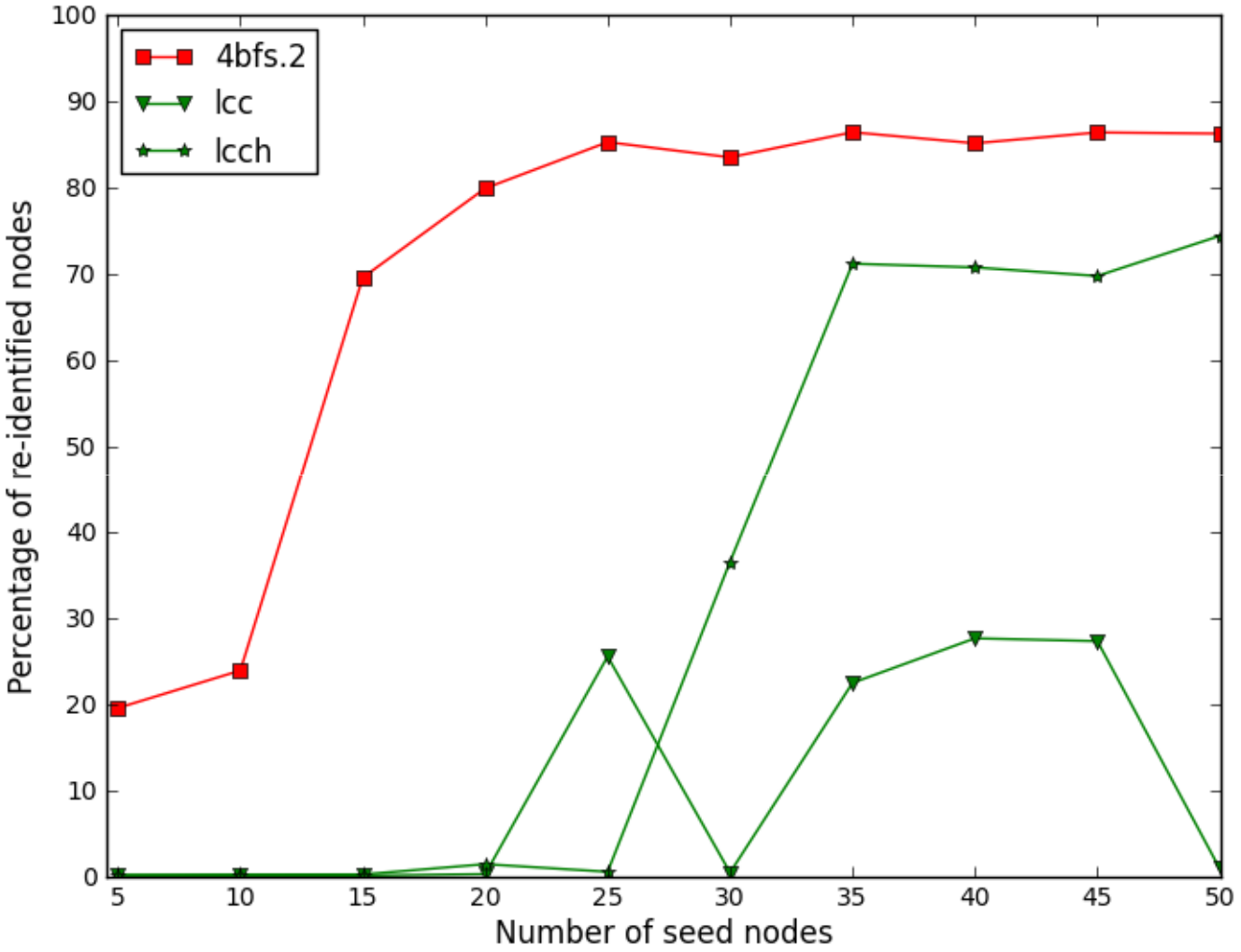



How node degree effects results:

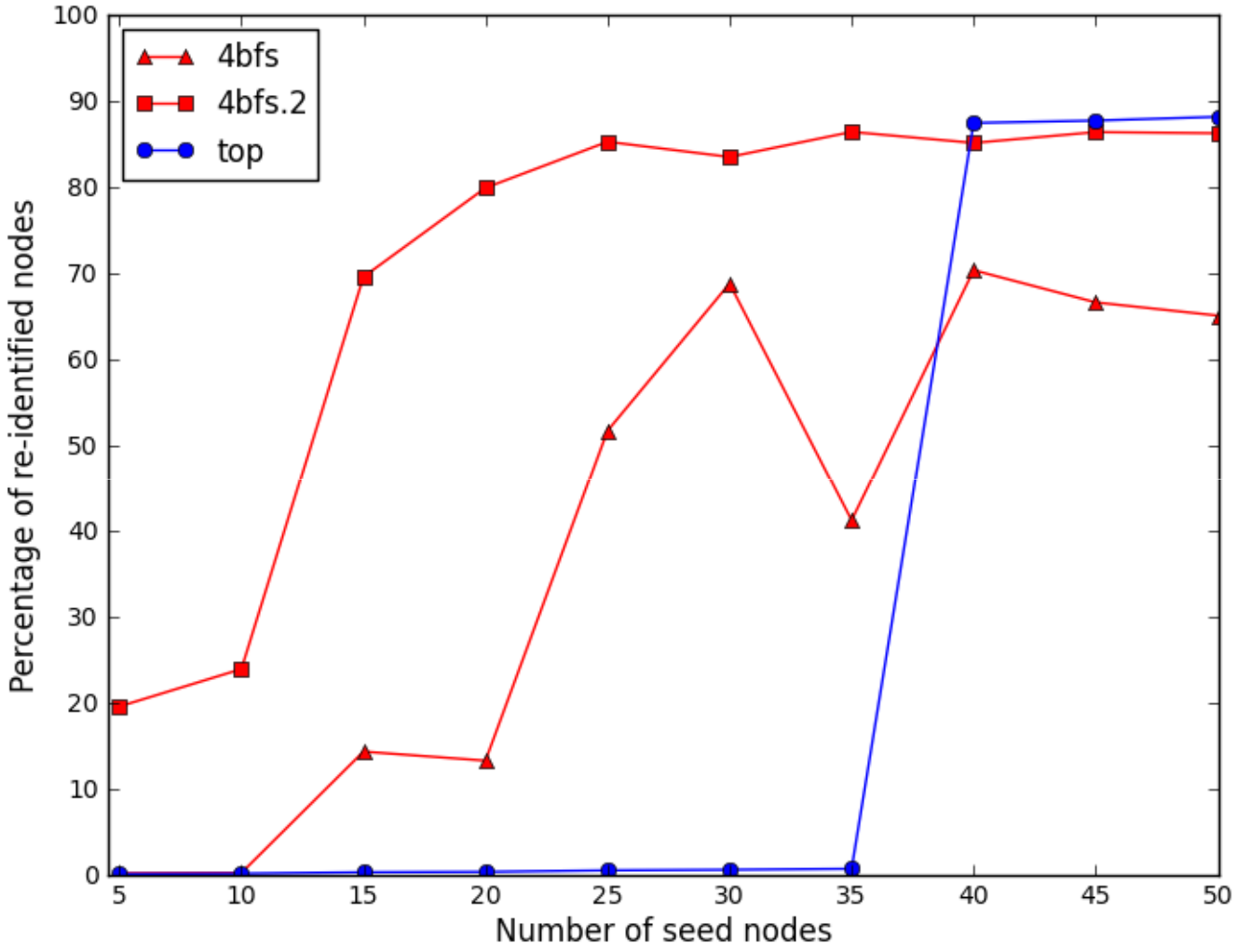
- Propagation limited (4bfs)
- Higher degree, less seeds

# Measurements of some properties (2)



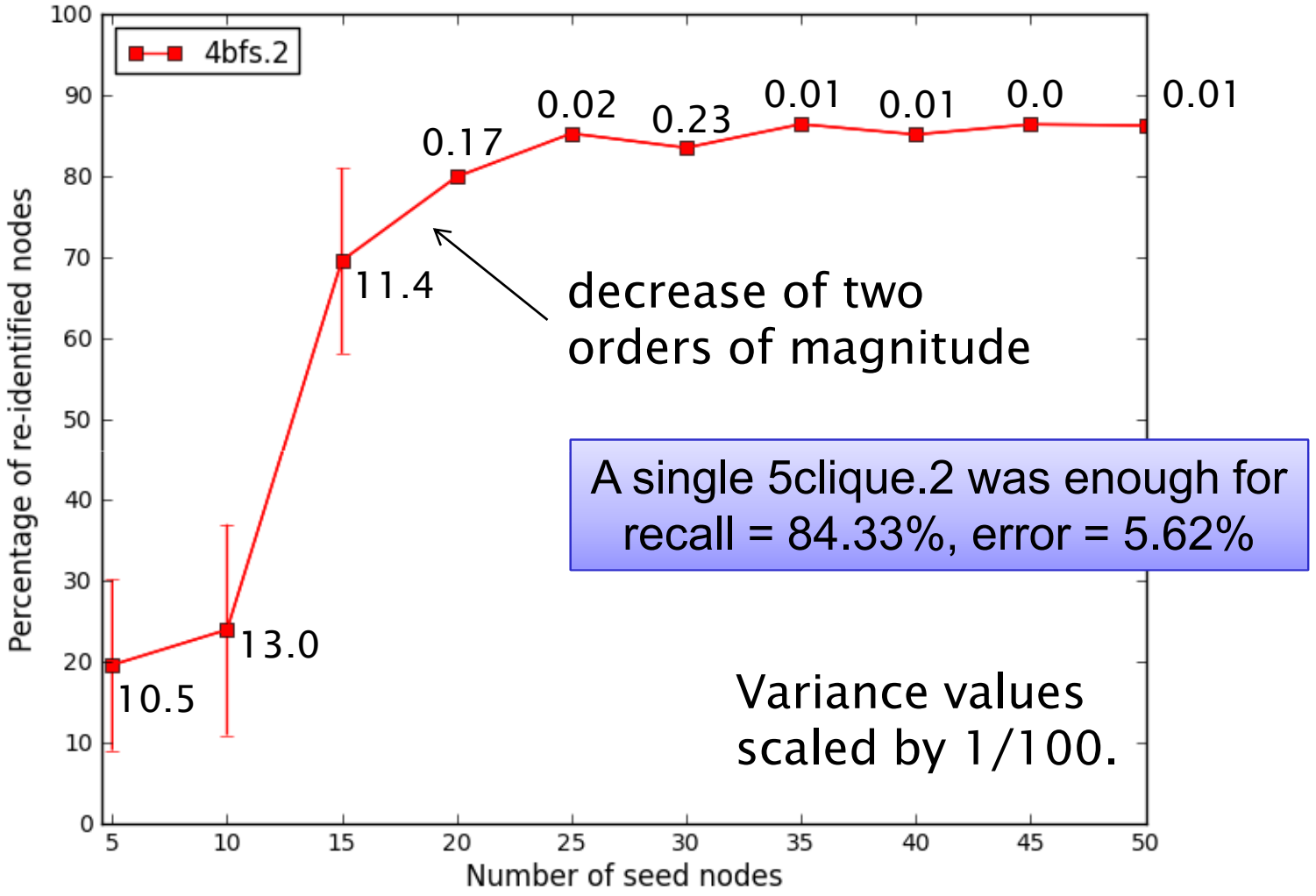

Not all methods are good for seeding

# Measurements of some properties (3)




Effects on phase transition boundaries

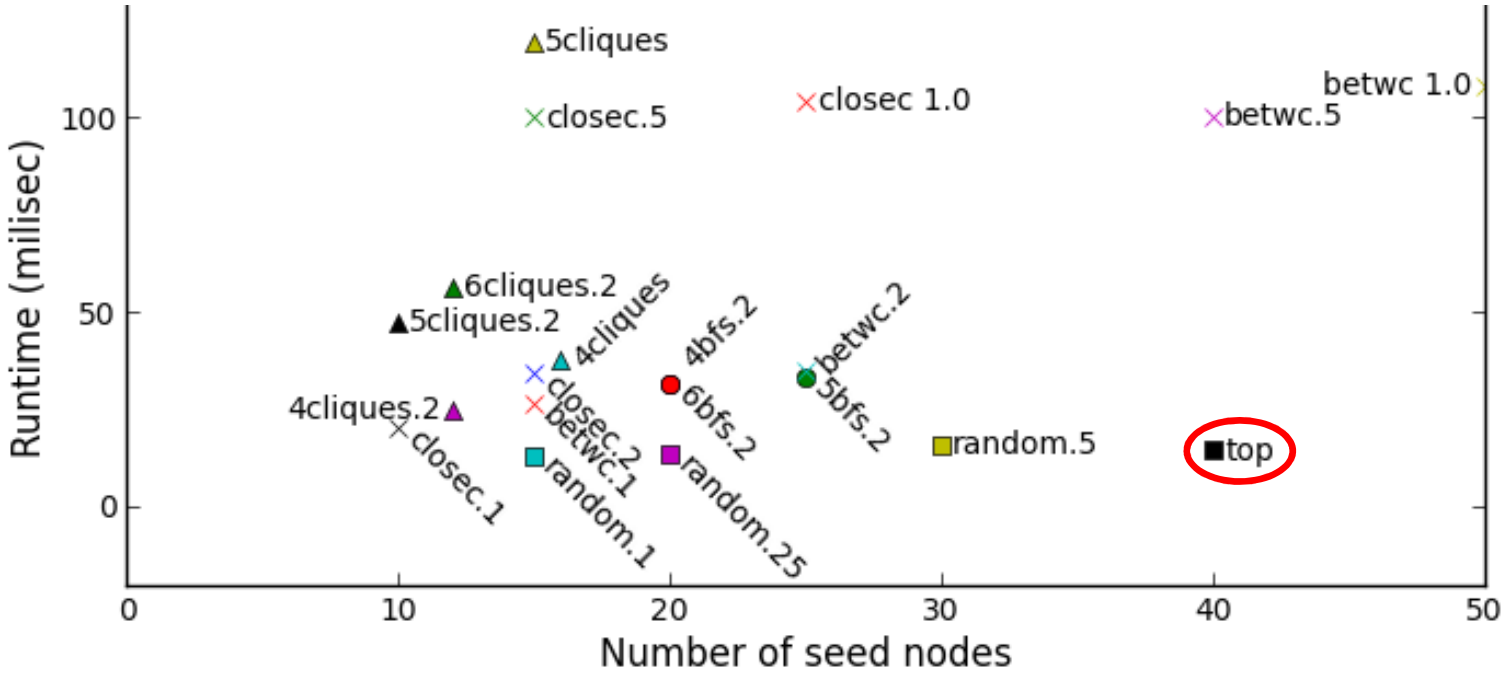
# Measurements of some properties (4)




Seeding stability illustrated

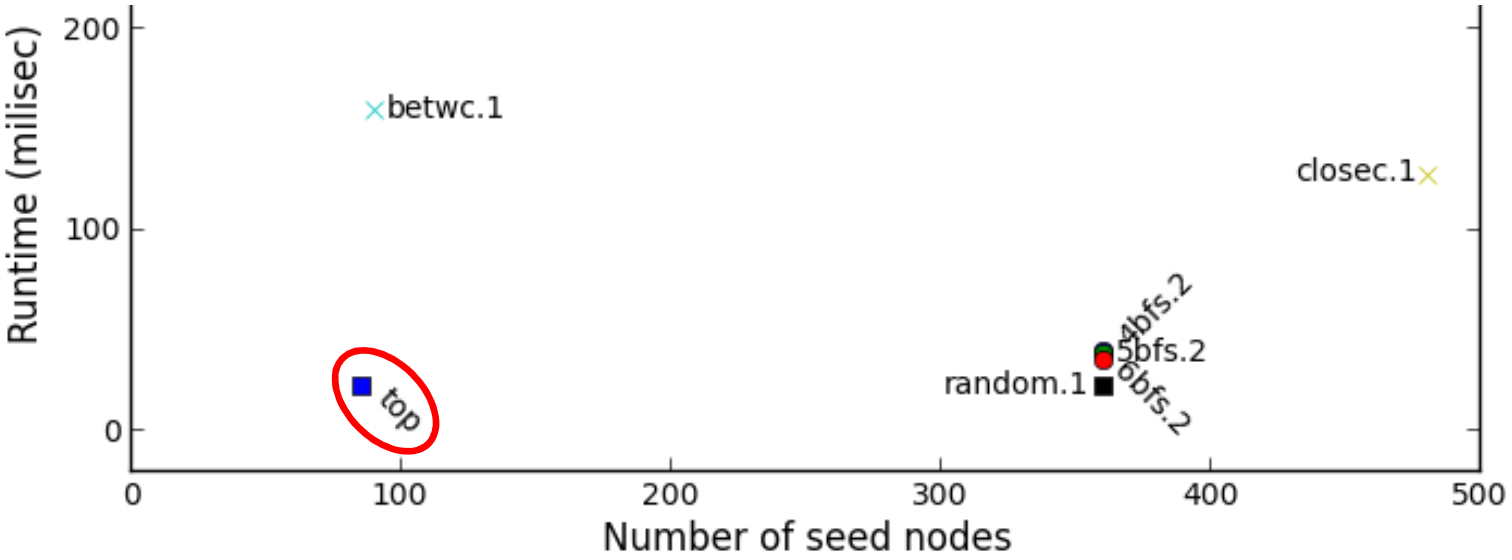
(1)

**LJ10k (small)**



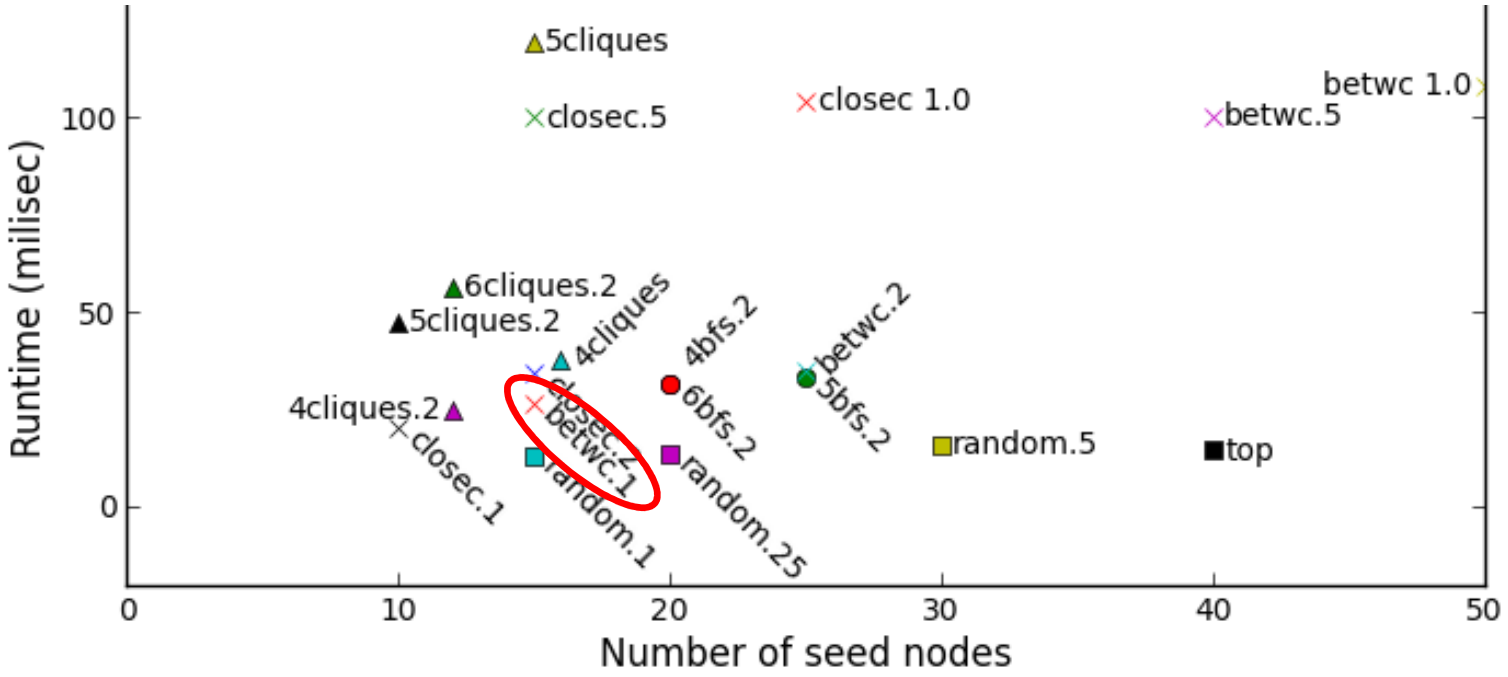

top: good results, but only in large networks

**Sd82k (large)**



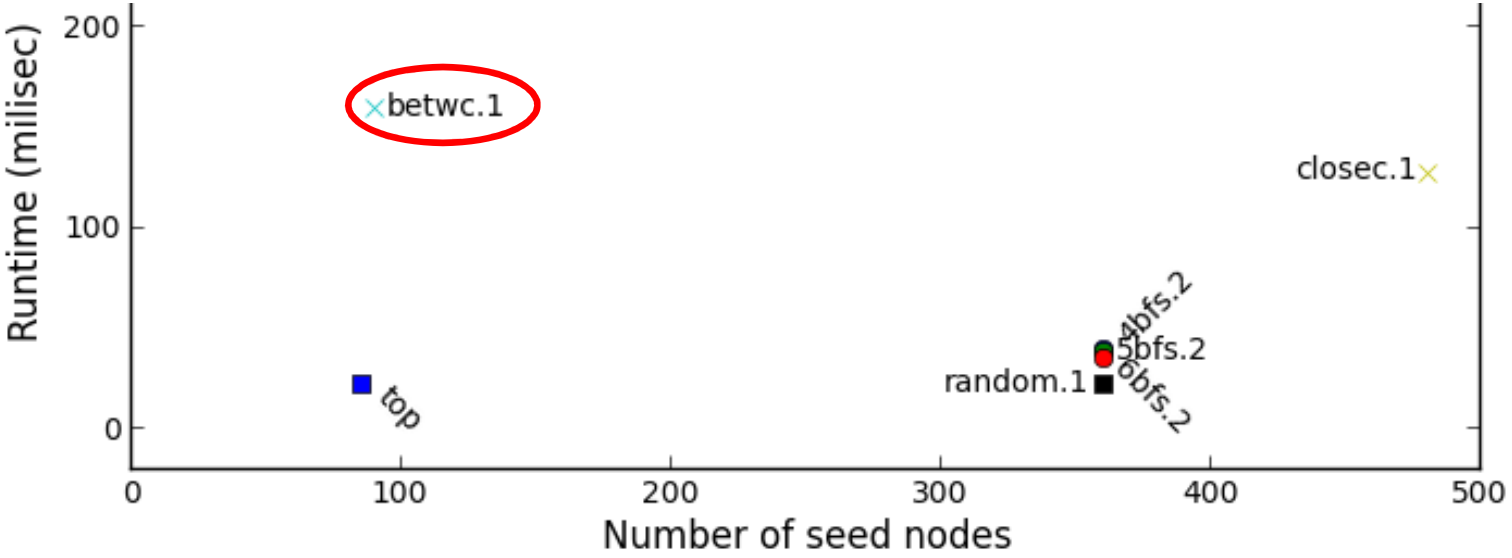
(2)

LJ10k (small)



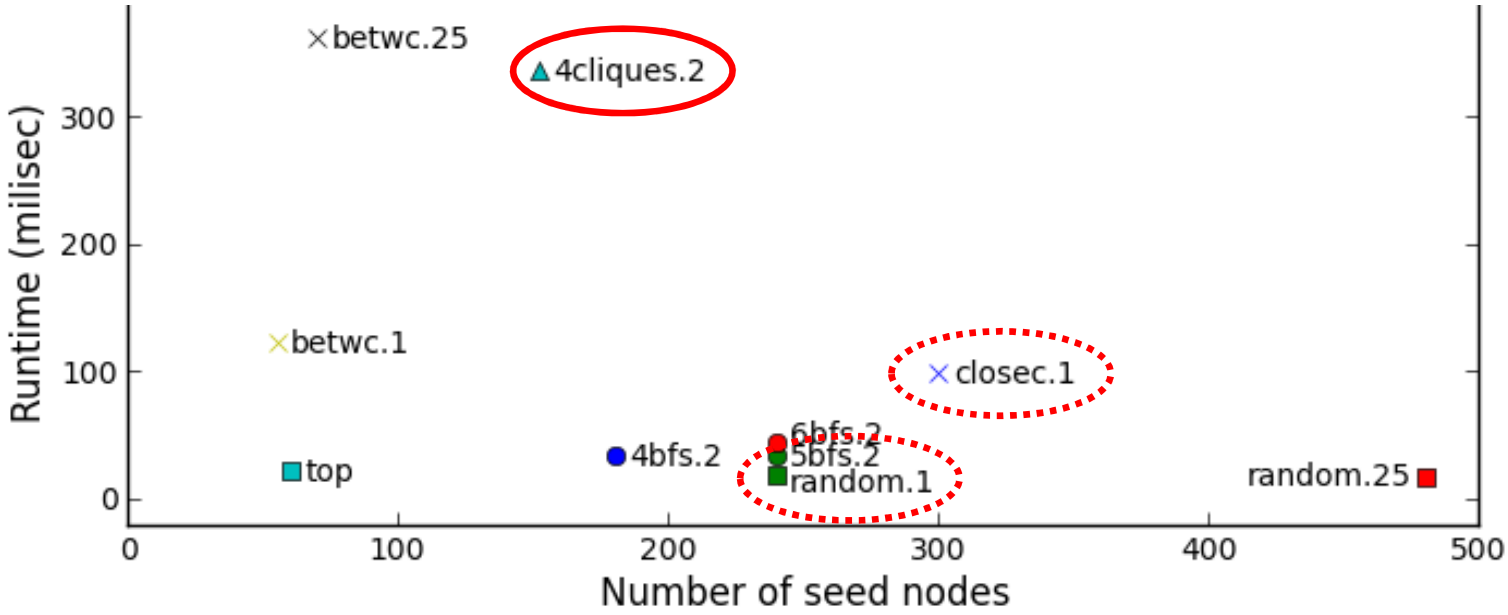

betwc.1: good choice regardless of network size!

Sd82k (large)



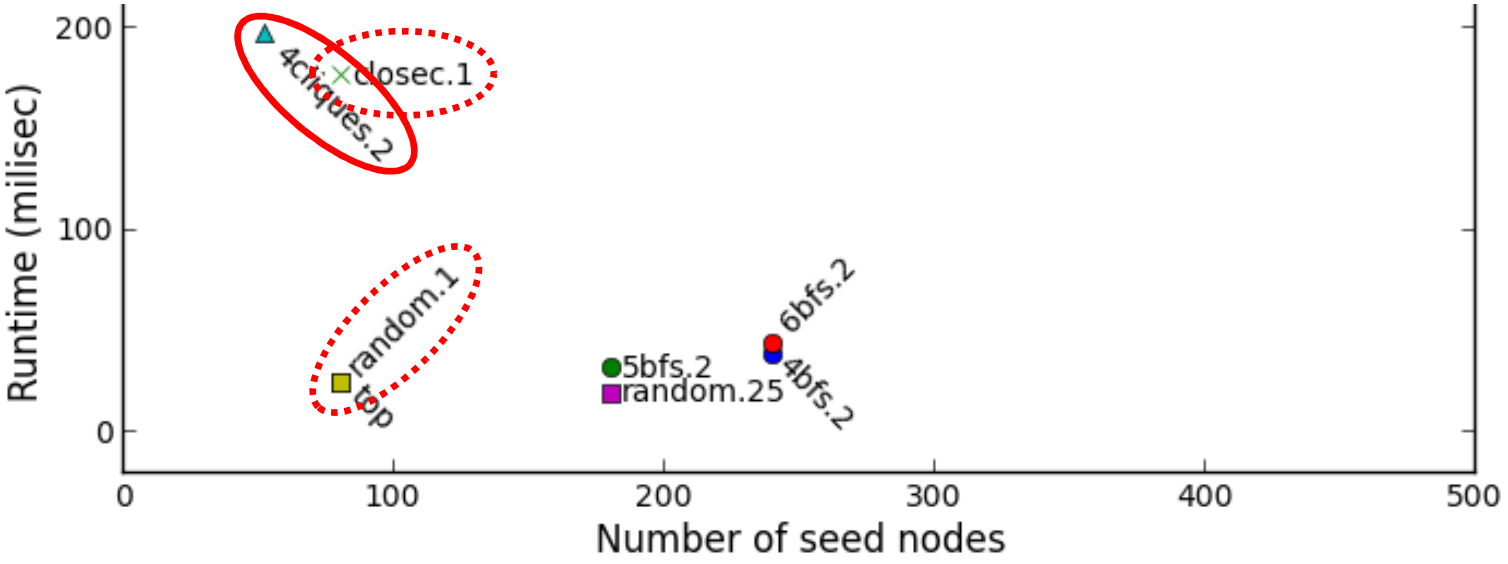
(3)

EP75k (sparse)

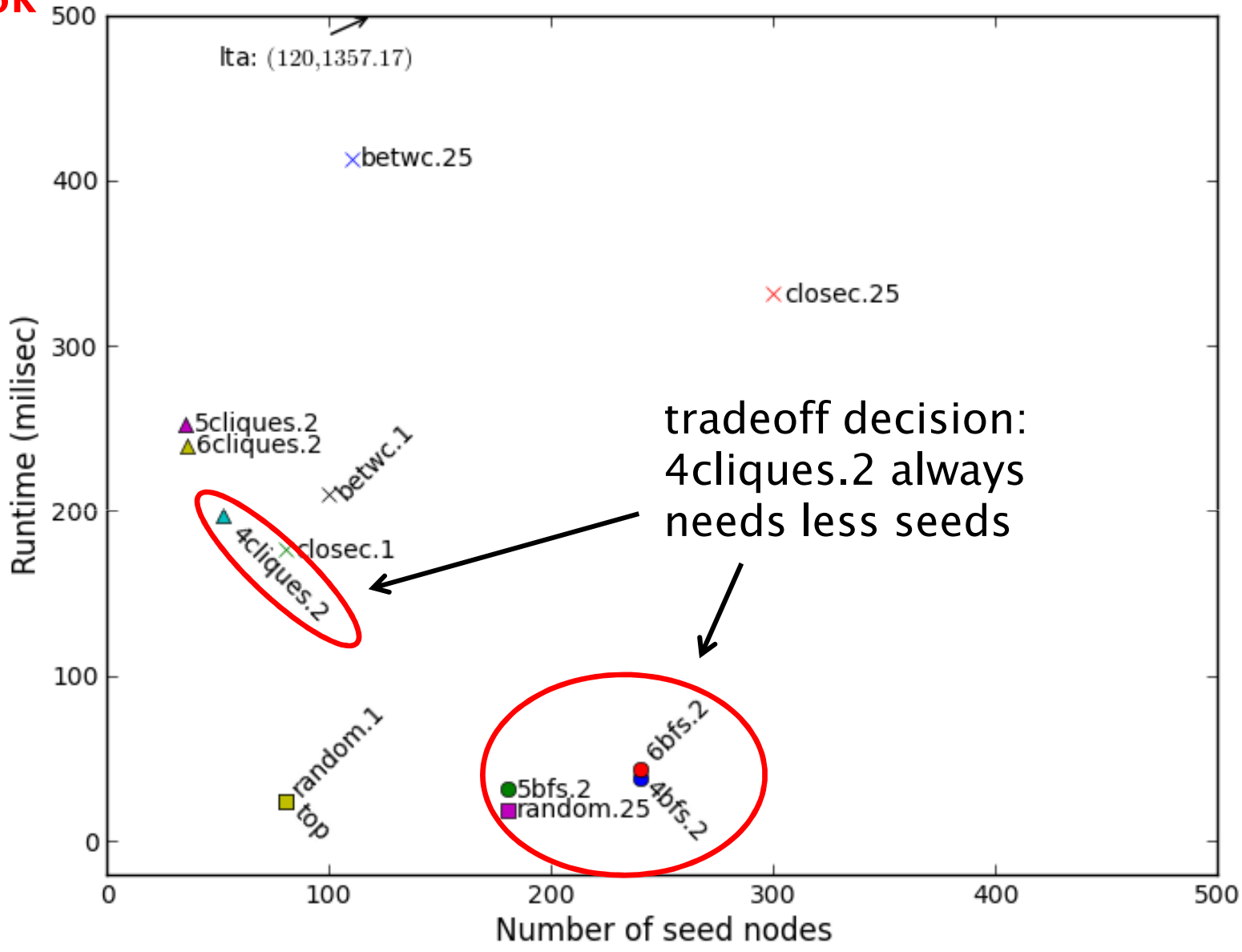



4cliques.2: good choice in dense networks
closec.1, random1.: seems OK also

Lj66k (dense)

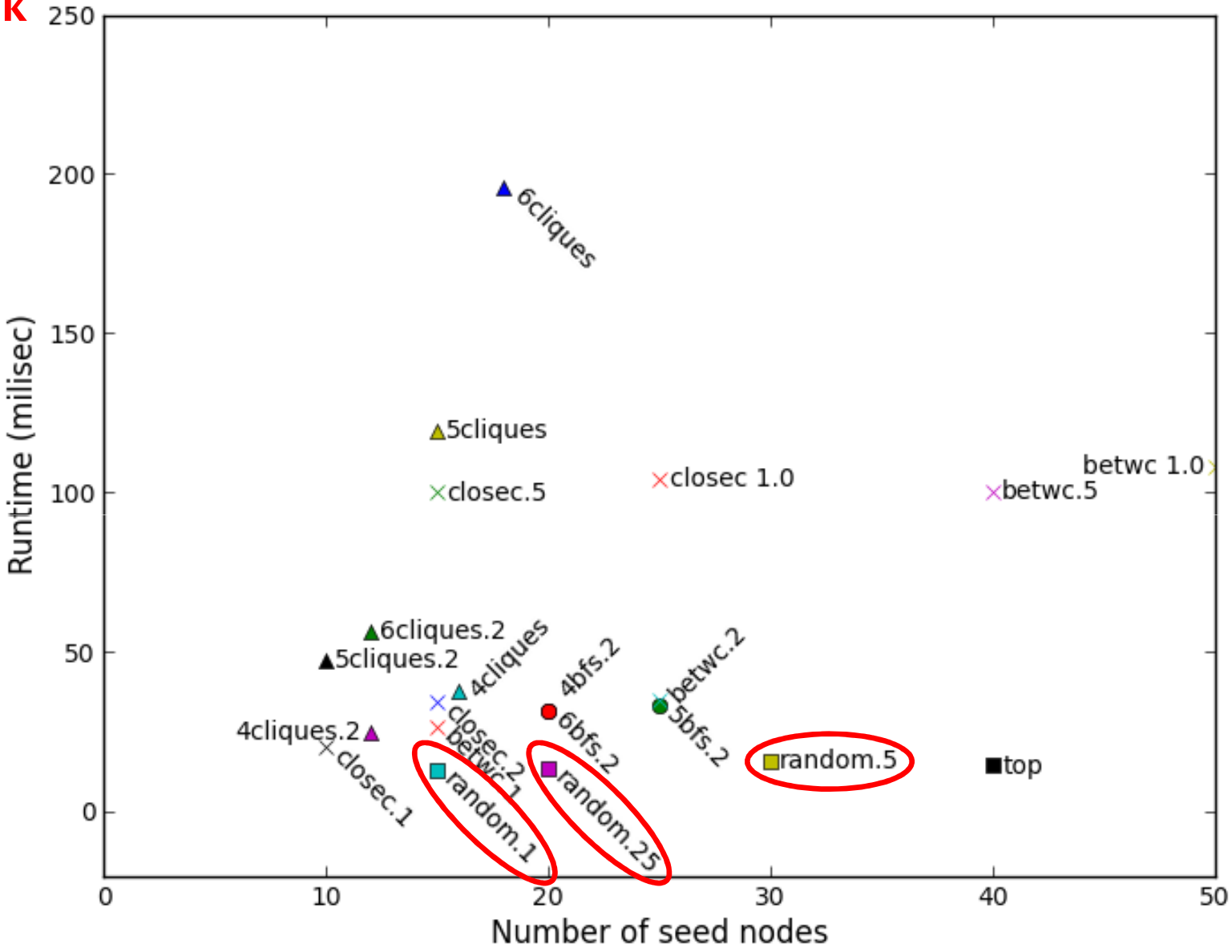


LJ66k




relationship of seed nodes matters

LJ10k




degree works as a good heuristic

# Conclusion & future work

- We emphasized that seeding can influence and limit propagation, for which one should consider
  - network size
  - network structure
- Properties demonstrated
  - global role of seed nodes (degree, betwc, closec)
  - local role of seed nodes (cliquish or BFS neighboring)
  - seed stability
  - phase transition dependency on seeding method (beside size, structure)
- We highlighted top performers
- Future work
  - Measure success of unstabe attacks
  - Updating attacker model in related work

# Thank you for your attention! Questions?

**Gábor György Gulyás**
gulyas@crysys.hu
Laboratory of Cryptography and System Security (CrySyS)
Budapest University of Technology and Economics
www.crysys.hu