

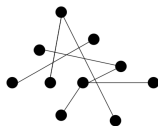
network *sampling*

introduction to *network analysis* (*ina*)

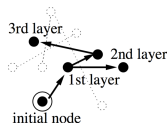
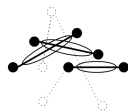
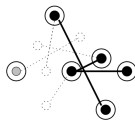
Lovro Šubelj  
University of Ljubljana  
spring 2021/22

# sampling *overview*

- *snowball sampling* and *contact tracing* in *sociology*  
network-based methods for sampling hidden populations
- *fractality* and *self-similarity* of *real networks* [SHM05]  
fractality is property of object that it is similar to part of itself  
self-similarity demands power-law size scaling under renormalization
- any *real network* is just *sample* of *true network* [BŠWB15]



original network



node/link selection and snowball sampling

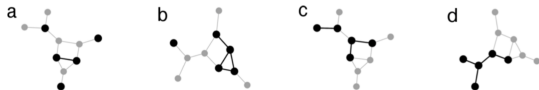
## sampling *populations*

- *snowball sampling* similar to *breadth-first search*  
nodes sampled proportional to their eigenvector centrality
- *contact tracing* similar to biased *best-first search*  
nodes sampled according to some hidden variable
- *respondent-driven sampling* similar to *random walk*  
nodes sampled proportional to their degree

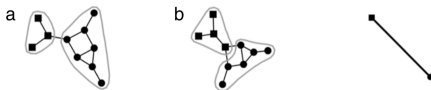
$$\text{estimate } \langle x \rangle = \frac{\sum_i x_i}{\sum_i 1} \quad \text{corrected } \langle x \rangle = \frac{\sum_i x_i / k_i}{\sum_i 1 / k_i}$$

# sampling *methods*

- *random selection methods* for *global network sparsification*  
random node/link selection w/ or w/o induction etc.
- *network exploration methods* for *local network sampling*  
random walks, snowball and expansion sampling etc.
- *merging/aggregation methods* for *network simplification*  
box covering, cluster growing, community aggregation etc.



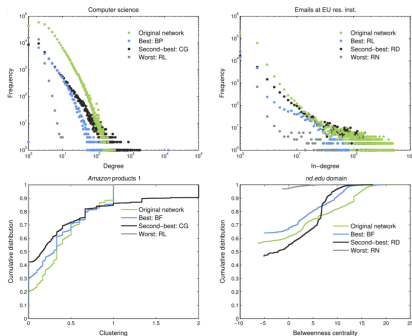
(a) node selection (b) by degree (c) link selection and (d) snowball sampling



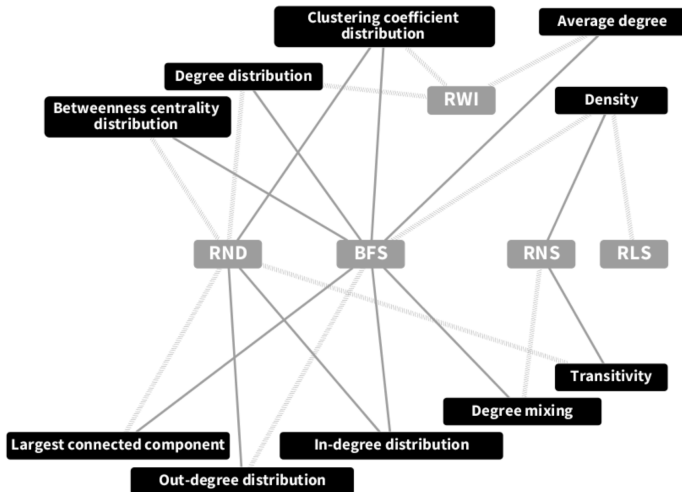
(a) community aggregation and (b) cluster growing

# sampling *networks*

- *similar sampled networks* for  $\gtrsim 15\%$  *network* [LF06]
  - *methods indistinguishable* for  $\ll 1\%$  *network* [BŠB17]
  - *selection/exploration*  $\ggg$  *aggregation methods* [BŠB14]
- independent of network type or size but not sample size

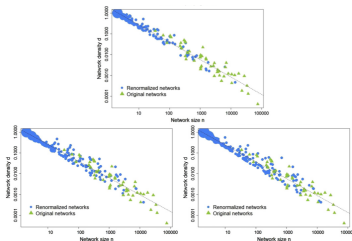


# sampling *scheme*

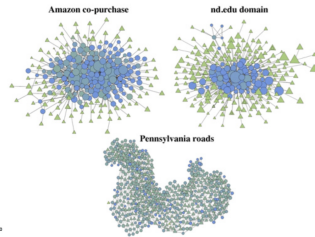


# sampling *density*

- *self-similar network density* of *real networks* [LJT<sup>+</sup>11]  
most real networks are sparse with  $\rho \approx 7.89n^{-0.99}$
- *network density invariant* under *any renormalization* [BŠB12]  
box covering, cluster growing, community aggregation etc.



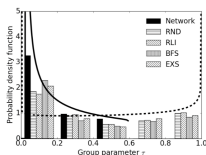
density invariance under renormalization



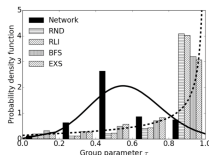
renormalized networks

# sampling *communities*

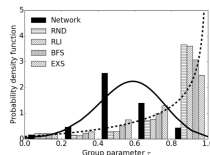
- *pronounced community structure* in *real networks* [GN02]  
many real networks contain communities with  $\tau \gg 0$
- *community structure enhanced* under *any sampling* [BŠWB15]  
random selection, snowball and expansion sampling etc.



citation network



social network



collaboration network





# sampling *references*



A.-L. Barabási.

*Network Science.*

Cambridge University Press, Cambridge, 2016.



Neli Blagus, Lovro Šubelj, and Marko Bajec.

Self-similar scaling of density in complex real-world networks.

*Physica A*, 391(8):2794–2802, 2012.



Neli Blagus, Lovro Šubelj, and Marko Bajec.

Assessing the effectiveness of real-world network simplification.

*Physica A*, 413:134–146, 2014.



Neli Blagus, Lovro Šubelj, and Marko Bajec.

Empirical comparison of network sampling: How to choose the most appropriate method?

*Physica A*, 477:136–148, 2017.



Neli Blagus, Lovro Šubelj, Gregor Weiss, and Marko Bajec.

Sampling promotes community structure in social and information networks.

*Physica A*, 432:206–215, 2015.



Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj.

*Exploratory Social Network Analysis with Pajek: Expanded and Revised Second Edition.*

Cambridge University Press, Cambridge, 2011.



David Easley and Jon Kleinberg.

*Networks, Crowds, and Markets: Reasoning About a Highly Connected World.*

Cambridge University Press, Cambridge, 2010.



Ernesto Estrada and Philip A. Knight.

*A First Course in Network Theory.*

Oxford University Press, 2015.

# sampling *references*



M. Girvan and M. E. J Newman.  
Community structure in social and biological networks.  
*P. Natl. Acad. Sci. USA*, 99(12):7821–7826, 2002.



Jure Leskovec and Christos Faloutsos.  
Sampling from large graphs.  
In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 631–636, Philadelphia, PA, USA, 2006.



Paul J. Laurienti, Karen E. Joyce, Qawi K. Telesford, Jonathan H. Burdette, and Satoru Hayasaka.  
Universal fractal scaling of self-organized networks.  
*Physica A*, 390(20):3608–3613, 2011.



Mark E. J. Newman.  
*Networks*.  
Oxford University Press, Oxford, 2nd edition, 2018.



Chaoming Song, Shlomo Havlin, and Hernan A. Makse.  
Self-similarity of complex networks.  
*Nature*, 433(7024):392–395, 2005.