Star Sampling and its Application to Online Social Networks

Maciej Kurant, Minas Gjoka, Carter Butts, Athina Markopoulou
Networking Group @ UC Irvine

Sampling OSNs

Our previous work


M. Gjoka, M. Kurant, C. Butts and A. Markopoulou, “Finding these Networks on Sampling OSNs (un) by Tracing the Board Game,” JSAC 2011.

E.g., Towards Unbiased BFS (Breadth-First Search) Sampling:

- Graph transistions on RG(μg)
- Observation
- Not all categories are equal
- Stratification. Node weight is proportional to its sampling probability under induced subgraph and star sampling.

Stratified WRW

- Measurement objective
- Category weights optimal under WIS
- Problem: Poor or no connectivity

Resolution: Limit the weight of tiny irrelevant categories.

- Modified category weights

- Edge weights in G

Measurement objective

- WRW sample

Induced Subgraph Sampling

- Star Sampling

- Final result

Often available when sample is collected through html scraping.

Our current interest: Node categories

E.g., country/city, workplace, school, first name, favorite music type, language

Coarse-grained topology

w(A, B) = Pr{a random node in A and a random node in B are connected}

We estimate w(A, B) based on a uniform on non-uniform sample of nodes, under induced subgraph and star sampling.

www.geosocialmap.com

(Random world according to Facebook)

Public and private colleges in the USA

Friendly countries in the Middle East

M. Kurant, M. Gjoka, C. Butts, and A. Markopoulou. “Walking on a Graph with a Magnifying Glass.” SIGMETRICS 2011


Colleges in Facebook

- 5-SRR collects 10-100 times more samples per college than RW
- This difference is larger for small colleges – identification works!