Star Sampling and its Application to Online Social Networks

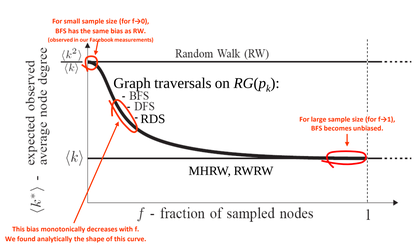
Maciej Kurant, Minas Gjoka, Carter Butts, Athina Markopoulou
Networking Group @ UC Irvine

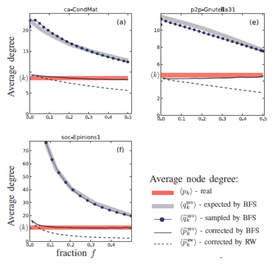
Sampling OSNs

Our previous work

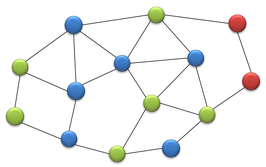
M. Gjoka, M. Kurant, C. T. Butts and A. Markopoulou, "Walking in Facebook: A Case Study of Unbiased Sampling of OSNs", *INFOCOM 2010.*
M. Kurant, A. Markopoulou and P. Thiran, "Towards Unbiased BFS Sampling", *JSAC* on Measurement of Internet Topologies 2011 (also ITC 2010)
M. Gjoka, C. T. Butts, M. Kurant and A. Markopoulou, "Multigraph Sampling of Online Social Networks", *JSAC 2011*
M. Gjoka, M. Kurant, C. Butts and A. Markopoulou, "Practical Recommendations on Sampling OSN Users by Crawling the Social Graph.", *JSAC 2011*

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



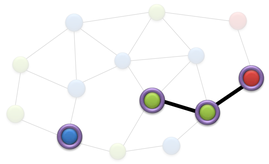


Our current interest: Node categories



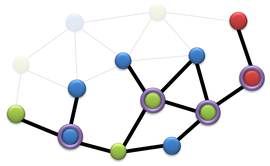
E.g., country/city, workplace, school, first name, favorite music type, language

Induced Subgraph Sampling



Sampled node
Observed edge
(known categories of both end-nodes)

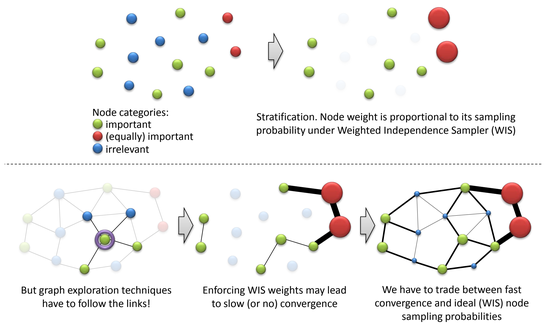
Star Sampling

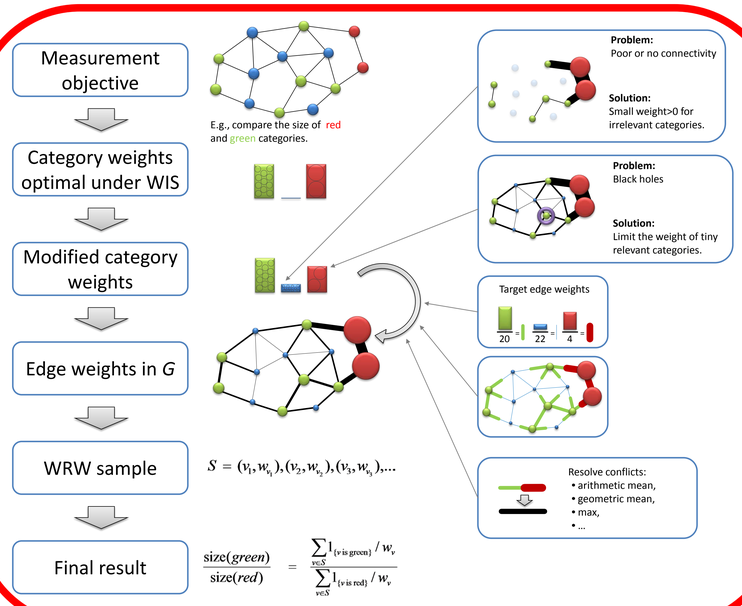


Often available when sample is collected through html scraping.

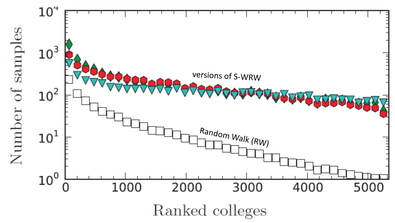
Stratified WRW

Not all categories are equal





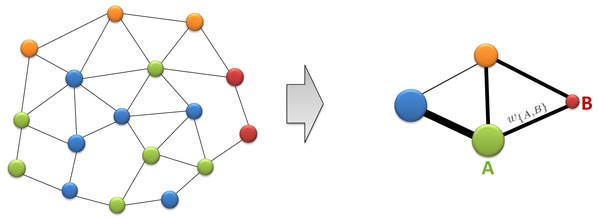
Colleges in Facebook



- S-WRW collects 10-100 times more samples per college than RW
- This difference is larger for small colleges – stratification works!

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

Coarse-grained topology



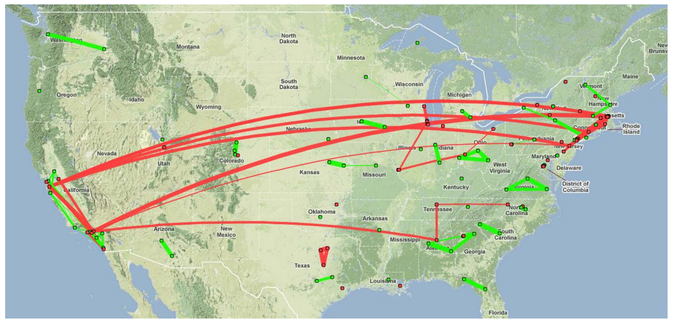
w(A,B) = Pr[a random node in **A** and a random node in **B** are connected]

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

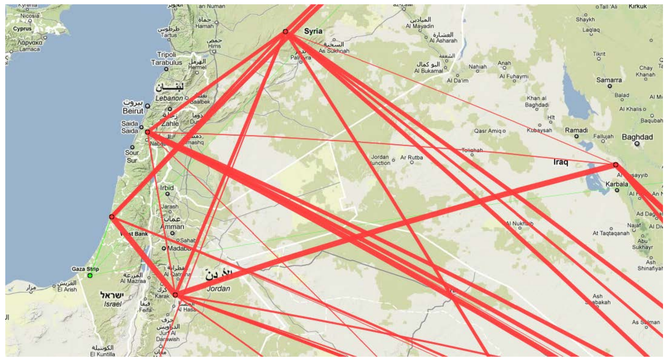
www.geosocialmap.com

(The world according to Facebook)

Public and private colleges in the USA



Friendly countries in the Middle East



M. Kurant, M. Gjoka, Y. Wang, Z. Almquist, C. Butts, and A. Markopoulou, "Coarse-Grained Topology Estimation via Graph Sampling," WOSN '12.