Resampling for Network Data

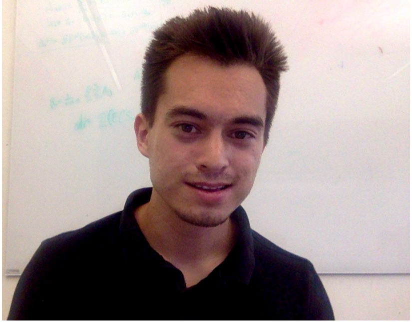
Purnamrita Sarkar

University of Texas at Austin

Joint work with Robert Lunde and Qiaohui Lin

October 22, 2020

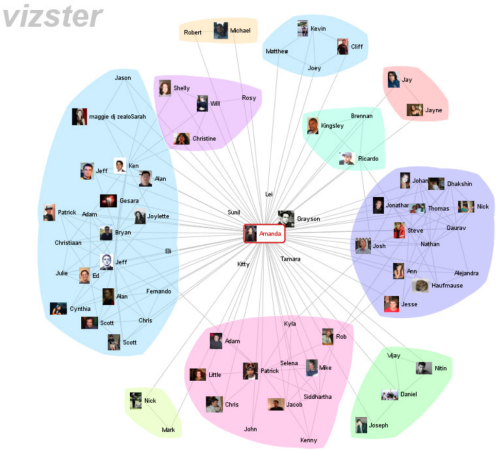
Co-authors



Robert Lunde

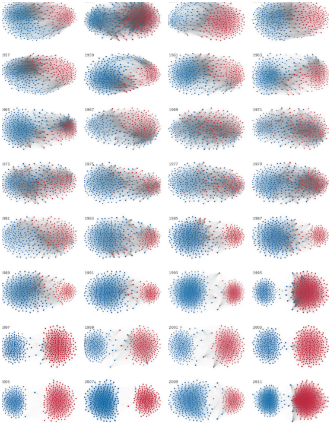
Qiaohui Lin

Networks Everywhere



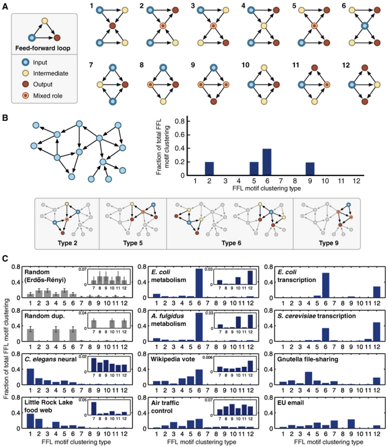
The Friendster Social Network (Heer and Boyd, 2005)

Networks Everywhere



Network of U.S. congress (Andris et al., 2015)

Networks Everywhere



Subgraphs/motifs in real networks (Gorochowski et al., 2018)

Motivation

- Inferential methods for network data are needed to address relevant scientific questions.
- In other settings, resampling methods allow valid inferences in a wide range of situations.
- Under the sparse graphon model, we study network analogs of
  - Subsampling
  - Jackknife

What is a graphon?

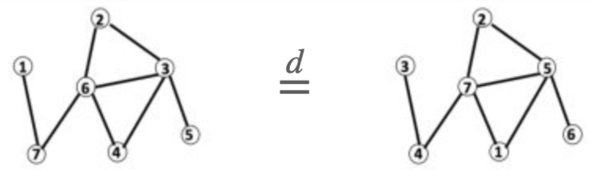


Figure from Crane (2018)

- Vertex exchangeability is a natural conditon to impose on a graph: asserts that distribution of random graph is unchanged when node labels are permuted.
- Representation theorems (Aldous-Hoover Theorem) assert that any binary (infinite-dimensional) array with the vertex exchangeability property has a certain form.

What is a Graphon, Cont.

- Aldous-Hoover Theorem asserts that any binary (infinite-dimensional) array with the vertex exchangeability property may be represented as mixture of processes with the following form:

$$A_{ij} \sim \text{Bernoulli}(w(\xi_i, \xi_j))$$

where $w : [0,1]^2 \mapsto [0,1]$, $\xi_1, \xi_2, \ldots \sim \text{Uniform}[0,1]$.
- The function w is the graphon (graphon function)

The Sparsity Problem with Graphons

- Consider the model:

$$A_{ij} \sim \text{Bernoulli}(w(\xi_i, \xi_j))$$

  where $w : [0,1]^2 \mapsto [0,1]$, $\xi_1, \xi_2, \ldots \sim \text{Uniform}[0,1]$.
- One major issue with this model is that the expected number of edges for $n \times n$ adjacency matrix is $O(n^2)$
- However, real world graphs are much sparser!

The Sparse Graphon Model

- Consider the following alternative model: Let $\{A^{(n)}\}_{n\geq 1}$ be a sequence of adjacency matrices of the form:

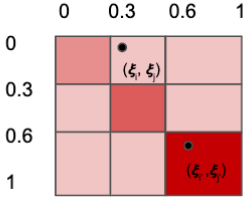
$$A_{ij}{}^{(n)} \sim \text{Bernoulli}(\rho_n w(\xi_i, \xi_j) \wedge 1)$$

where $w : [0,1]^2 \mapsto \mathbb{R}$,
$\xi_1, \xi_2, \ldots \xi_n \sim \text{Uniform}[0,1]$, $\rho_n \to 0$, $\int_0^1 \int_0^1 w(u,v)du\ dv = 1$.

- The sequence ρ_n (unknown) controls sparsity level

Sparse Graphon Model: Example 1

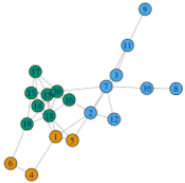
- Three-communities Stochastic Block Model Holland et al. (1983)
  - Partition vertex set $\{1,...,n\}$ into three disjoint communities $\{C_1, C_2, C_3\}$ with membership probability $(0.3, 0.3, 0.4)$
  - Community-Community Interaction Matrix



Discretize uniform ξ

	C1	C2	C3
C1	0.4	0.1	0.1
C2	0.1	0.5	0.1
C3	0.1	0.1	0.7

Community Interaction Matrix

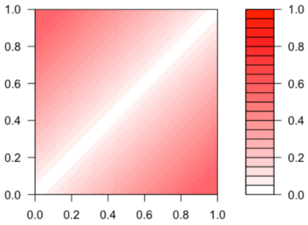


SBM n=20, $\rho_n = 1$

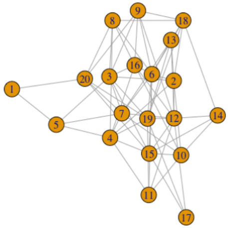
Sparse Graphon Model: Example 2

- Continuous Graphon Model

$$h_n(u,v) = P(A_{ij} = 1 \mid \xi_i = u, \xi_j = v) = \rho_n |u - v| \quad \text{(GR2)}$$



Graphon $w(\xi_i, \xi_j) = |\xi_i - \xi_j|$



GR(2) n=20, $\rho_n = 1$

Some Intuition About Sparse Graphons

- We still have a lot of structure with sparse graphon models.
  - $A^{(n)}$ is a function of independent random variables.
  - $A_{ij}^{(n)}$ and $A_{kl}^{(n)}$ are generally dependent, but independent if $\{i,j\} \cap \{k,l\} = \emptyset$.
- Let $P_{ij}^{(n)} = w(\xi_i, \xi_j)$. We can view $A_{ij}^{(n)}/\rho_n = P_{ij}^{(n)} + \epsilon_{ij}$, where ϵ_{ij} are independent conditional on $\boldsymbol{\xi}_n$.
- For certain (important) functions, contribution of ϵ_{ij} will be negligible as $n \to \infty$, $\rho_n \to 0$ sufficiently fast; $P_{ij}^{(n)}$ is often the "signal."

The Target Parameter

- We will look to infer properties of the underlying graphon w.
- To conduct statistical inference (confidence intervals, hypothesis tests), we will look to characterize limiting distribution of estimators centered at this parameter.
- Examples:
    - Limiting triangle frequency:

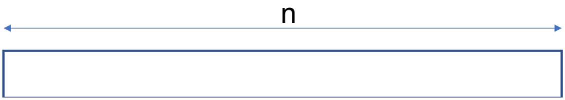
$$\int_0^1 \int_0^1 \int_0^1 w(x,y)w(y,z)w(z,x)\ dx\ dy\ dz$$

    - Eigenvalues:

$$\int_0^1 w(u,v)f_r(v)dv = \lambda_r(w)f_r$$

Subsampling

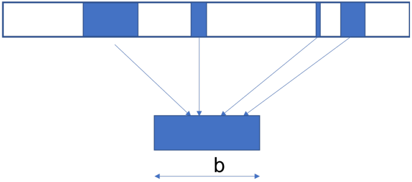
Subsampling: the IID Case



1. Inferential quantity of interest θ
2. Goal, want to estimate the distribution of $\hat{\theta}$
3. If I could draw N size n samples, and knew θ (this is silly, really) this will not be difficult, we will do

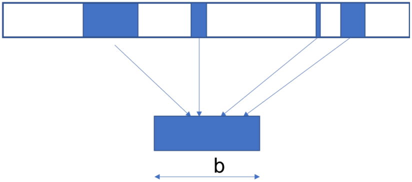
$$\frac{1}{N}\sum_{i=1}^{N} 1(\hat{\theta}_i - \theta \leq t)$$

Subsampling: the IID Case



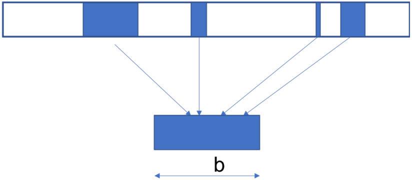
1. Since we have only one dataset, we pretend that is the population.
2. Now we pick N size b subsets without replacement.
3. Hope is this size b sample behaves like a size b sample from the underlying distribution.

Subsampling: the IID Case



1. So, using the plugin principle, can we just do: $\frac{1}{N}\sum_{i=1}^{N} 1(\hat{\theta}_b - \hat{\theta}_n \leq t)$

Subsampling: the IID Case



1. So, using the plugin principle, can we just do: $\frac{1}{N}\sum_{i=1}^{N} 1(\hat{\theta}_b - \hat{\theta}_n \leq t)$
2. Well, not quite, since your scale is smaller, and the variances of the estimators will be larger–so we need to correct for it. $\frac{1}{N}\sum_{i=1}^{N} 1(\tau_b(\hat{\theta}_b - \hat{\theta}_n) \leq t)$

Subsampling: the IID Case

- Works under weaker conditions than bootstrap (Politis and Romano, 1994).
- If $b_n \to \infty$ and functional converges in distribution, size b and size n functionals should be close.

Vertex Subsampling for Networks

- Suppose we sample b vertices w/o replacement from a size n graph, and take the induced subgraph ($b \times b$ submatrix of adjacency matrix $A^{(n,b)}$)
- For (dense) graphons, a similar principle applies.
- Size b adjacency matrix: $A^{(1)}, A^{(2)}, \dots A^{(b)}, \dots A^{(n)}$

$$A_{ij}^{(b)} \sim \mathrm{Bernoulli}(w(\xi_i, \xi_j))$$

- Induced subgraph: $A^{(1)}, A^{(2)}, \dots A^{(b)}, \dots A^{(n)}$

$$A_{ij}^{(n,b)} \sim \mathrm{Bernoulli}(w(\xi_i, \xi_j))$$

Vertex Subsampling for Networks, Continued

- Suppose we take b vertices from a size n graph, and take the induced subgraph ($b \times b$ submatrix of adjacency matrix $A^{(n,b)}$)
- For (sparse) graphons, induced subgraph is sparser .
- Size b adjacency matrix: $A^{(1)}, A^{(2)}, \ldots A^{(b)}, \ldots A^{(n)}$

$$A_{ij}^{(b)} \sim \mathrm{Bernoulli}(\rho_b w(\xi_i, \xi_j) \wedge 1)$$

- Induced subgraph: $A^{(1)}, A^{(2)}, \ldots A^{(b)}, \ldots A^{(n)}$

$$A_{ij}^{(n,b)} \sim \mathrm{Bernoulli}(\rho_n w(\xi_i, \xi_j) \wedge 1)$$

- However, if functional converges in distribution when size b graph normalized by ρ_n, similar principle applies.

Validity of Vertex Subsampling

Define the following quantities:

- τ_n: normalizing sequence (typically $\tau_n = \sqrt{n}$).
- $L_{n,b}(t) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\tau_b[\hat{\theta}_{n,b}(A^{(n,b)}) - \hat{\theta}_n(A^{(n)})] \leq t)$
- $J_{n,b}(t)$: CDF of functional calculated on induced subgraph
- $J_n(t)$: CDF of sampling distribution
- $J(t)$: limiting distribution

Theorem 1 (Consistency of Vertex Subsampling)

Suppose that $N \to \infty$, $b_n \to \infty$, $b_n = o(n)$, $J_n(t) \to J(t)$, *and* $J_{n,b}(t) \to J(t)$. *Then,*

$$|L_{n,b}(t) - J_n(t)| \xrightarrow{P} 0$$

Weak Convergence For Network Functionals?

- Bhattacharyya and Bickel (2015) previously established subsampling validity for counts, Bickel et al. (2011) established CLT for counts
- While our subsampling validity result is more general, what other network functionals converge in distribution?
- We establish a central limit theorem for eigenvalues of the adjacency matrix generated by a sparse graphon.
- Previously LLN known for spectra of dense graphons (Borgs et al., 2012), CLT for Erdos Renyi (Füredi and Komlós, 1981), but CLT not known for general (sparse) graphons.

Central Limit Theorem for Eigenvalues

- Recall the integral operator:

$$T_w f = \int_0^1 w(u,v) f(v) dv$$

satisfying $T_w f_r = \lambda_r(w) f_r$. Spectral decomposition gives $w(u,v) = \sum_{r=1}^{R} \lambda_r(w)\phi_r(u)\phi_r(v)$ where R is rank.

- Define the following functional:

$$Z_{n,r} = \sqrt{n}[\lambda_r(A/n\rho_n) - \lambda_r(w)]$$

Theorem 2 (Central Limit Theorem for Eigenvalues)

Suppose that $\|w(u,v)\|_\infty < \infty$[a], $\rho_n \to 0$, $\rho_n = \omega(1/\sqrt{n})$, *and* $R < \infty$. *Then* $(Z_{n,1}, \ldots Z_{n,R}) \rightsquigarrow (Z_{\infty,1}, \ldots Z_{\infty,R})$. *If eigenvalues are distinct, then limiting distribution is multivariate Normal.*

[a] boundedness of w can be relaxed but leads to stronger sparsity conditions

Intuition for CLT for Eigenvalues

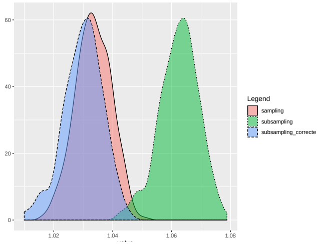
- For simplicity take $r = 1$ and drop the subscript.
- We can decompose Z_n as:

$$
\begin{aligned}
Z_n &= \sqrt{n}[\lambda_1(A/n\rho_n) - \lambda_1(w)] \\
&= \underbrace{\sqrt{n}\left[\frac{\lambda_1(A)}{n\rho_n} - \frac{\lambda_1(P)}{n}\right]}_{\text{Noise}} + \underbrace{\sqrt{n}\left[\frac{\lambda_1(P)}{n} - \lambda_1(w)\right]}_{\text{Signal}}
\end{aligned}
$$

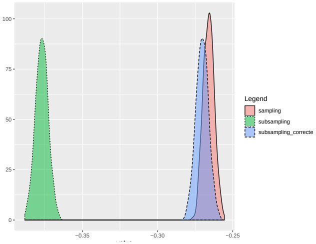
- Noise term concentrates due to results refined spectral perturbation results from Eldridge et al. (2018).
- CLT for signal component holds due to results on random matrix approximations of integral operators (Koltchinskii and Giné, 2001) .
- Unbounded graphons pose a bit more technical difficulty.

Simulation Study

Let $\nu_n = \rho_n \cdot P(A_{ij} = 1)$ and consider sparse stochastic block model: with $B = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/6 \end{pmatrix}$ and $\pi = (0.3, 0.3, 0.4)$. The corresponding graphon is rank 2 and has one positive and one negative eigenvalue, with $\lambda_1 = 1.035$ and $\lambda_2 = -0.267$.



(a)



(b)

Sampling and subsampling distributions for inference on $\lambda_1(w)$ and $\lambda_2(w)$.

Recap for Subsampling

- In IID settings, subsampling is known as a variant of bootstrap that is consistent under weaker conditions.
- We establish validity of subsampling for sparse graphons under similar conditions.
- We establish a CLT for eigenvalues, which yields subsampling validity.
- In practice, subsampling seems to estimate variance well for eigenvalues, but suffers from bias.

Jackknife

Network Count Functionals

- Let R denote the adjacency matrix of a subgraph of interest, with r vertices and s edges, with vertex set $V(R)$ and edge set $E(R) \subset V(Q) \times V(Q)$. $\overline{E(R)}$ is complement of $E(R)$. Normalized subgraph density is

$$\tilde{P}(R) = \rho_n^{-s} E\left[\prod_{(i,j)\in E(R)} \rho_n w(\xi_i, \xi_j) \prod_{(i,j)\in \overline{E(R)}} (1 - \rho_n w(\xi_i, \xi_j))\right]$$

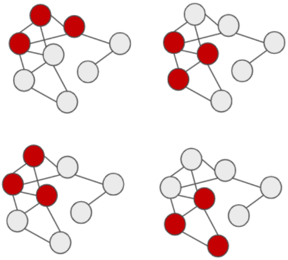
- Estimator for normalized subgraph density $\tilde{P}(R)$

$$\hat{P}(R) := \rho^{-s} \frac{1}{\binom{n}{r}} \sum_{1 \le i_1 < i_2 < \ldots < i_r \le n} \mathbb{1}(A^{(n)}_{i_1,\ldots,i_r} \cong R)$$

where $A^{(n)}_{i_1,\ldots,i_r} \cong R$ if there exists a permutation function π such that $A_{\pi(i_1),\ldots,\pi(i_r)} = R$.

Network Count Functionals

Example: Triangle Densities.
Examine every subset of 3 nodes, if they form a triangle.

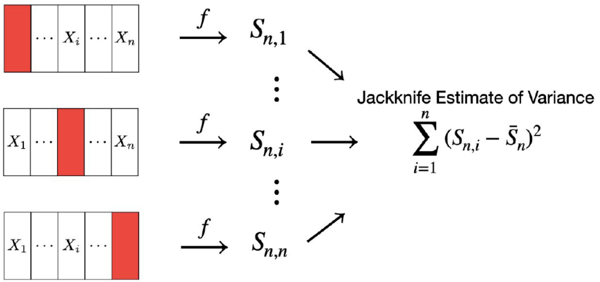


$$\text{Triangle density} := \frac{\sum_{i<j<k} A_{ij} A_{jk} A_{ik}}{\binom{n}{3} \rho_n^3} = \frac{4}{\binom{8}{3} \rho_n^3}$$

Jackknife and Efron-Stein Inequality on I.I.D Data

- Let $X_1, \dots X_n \sim P$, $S_n = f(X_1, \dots, X_n)$
- Let $S_{n,i}$ denote functional with X_i left out, $\bar{S}_n = \frac{1}{n}\sum_{i=1}^{n} S_{n,i}$. Then the Jackknife estimate for the variance $\mathrm{Var}\ S_{n-1}$ is

$$\widehat{\mathrm{Var}}_{JACK}\ S_{n-1} = \sum_{i=1}^{n}(S_{n,i} - \bar{S}_n)^2$$



Jackknife and Efron-Stein Inequality on I.I.D Data

- Jackknife is consistent for smooth functionals (see e.g. Shao and Tu (1995)), but it generally needs regularity conditions stronger than bootstrap.

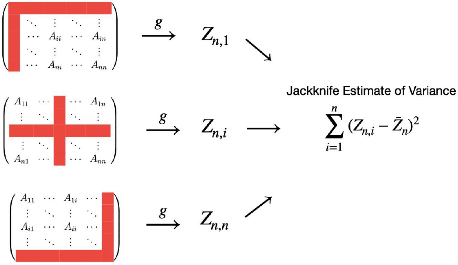
- Efron-Stein inequality (Efron (1979)):

$$\mathrm{Var}\ S_{n-1} \leq E(\widehat{\mathrm{Var}}_{JACK}\ S_{n-1})$$

Network Jackknife Procedure

- Network jackknife: leave-one-node out.
- Let $Z_{n,i}$ denote the r.v. by applying network functional g to the graph with node i removed and $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_{n,i}$.
- Jackknife estimate under sparse graphon setting is

$$\widehat{\mathrm{Var}}_{\mathrm{JACK}}\ Z_{n-1} := \sum_{i=1}^{n}(Z_{n,i} - \bar{Z}_n)^2$$



Network Efron-Stein Inequality

Theorem 3 (Network Efron-Stein Inequality)

For any functional invariant to node-permutation,

$$\mathrm{Var}\ Z_{n-1} \leq E(\widehat{\mathrm{Var}}_{JACK}\ Z_{n-1})$$

To prove this, we use martingale difference techniques (Rhee and Talagrand (1986)) with appropriate filtration (Borgs et al. (2008))

Consistency for Count Functionals

Theorem 4 (Consistency for Counts)

Suppose R is acyclic or a p-cycle. Then if $n\rho_n \to \infty$,

$$n \cdot \widehat{\mathrm{Var}}_{JACK} \; \hat{P}(R) \xrightarrow{P} \sigma^2$$

where $\sigma^2 = \lim_{n\to\infty} n \cdot \mathrm{Var} \; \hat{P}(R)$.

Consistency for Count Functionals

Proof Sketch:

$$\hat{P}(R) := \rho^{-s} \frac{1}{\binom{n}{r}} \sum_{1 \le i_1 < i_2 < \ldots < i_r \le n} \mathbb{1}(A^{(n)}_{i_1,\ldots,i_r} \cong R)$$

- $E[\hat{P}(R)|\xi_1, \ldots, \xi_n]$ is a U-statistic (unobserved)
- $\hat{P}(R) =$ U-statistic (Signal) $+$ Bernoulli perturbations (Noise)
- Jackknife Variance is consistent for U-statistics (Lee (1990)).
- Variance from the noise part is proved by us to be $o(\frac{1}{n})$ and negligible.

Consistency for Smooth Functions of Count Functionals

Let $f(G_n)$ denote a function of the vector $(\hat{P}(R_1), \ldots, \hat{P}(R_d))$.
Example: Normalized transitivity $:= \frac{\text{Triangle density}}{\text{Two star density}}$.

Theorem 5 (Informal - Consistency for Smooth Functions of Count Functional)

Under a broad set of conditions of f and its gradient and integrability of graphon, let σ_f^2 denote the asymptotic variance of $\sqrt{n}[f(G_n) - f(E(G_n))]$. Then,

$$n \cdot \widehat{\mathrm{Var}}_{JACK}\ f(G_n) \xrightarrow{P} \sigma_f^2$$

See full paper for formal statement of the theorem and detailed conditions.

Network Jackknife: Simulation Study

- Graphons Used
  - Stochastic Block Model (SBM)
    $B = ((0.4, 0.1, 0.1), (0.1, 0.5, 0.1), (0.1, 0.1, 0.7))$ and community membership probability $(0.3, 0.3, 0.4)$, and sparsity parameter $\rho_n = n^{-1/3}$
  - Continuous Graphon: GR2 with $\nu_n = n^{-1/3}$

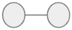
$$h_n(u, v) = P(A_{ij} = 1 \mid \xi_i = u, \xi_j = v) = \nu_n |u - v| \quad \text{(GR2)}$$

- Graph size $n = 100, 500, 1000, 2000, 3000$, each simulated 100 graphs.

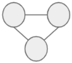
Network Jackknife: Simulation Study

- Count Functionals Used

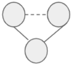
$$\text{Edge density} := \frac{\sum_{i<j} A_{ij}}{\binom{n}{2}\rho_n}$$



$$\text{Triangle density} := \frac{\sum_{i<j<k} A_{ij}A_{jk}A_{ki}}{\binom{n}{3}\rho_n^3}$$



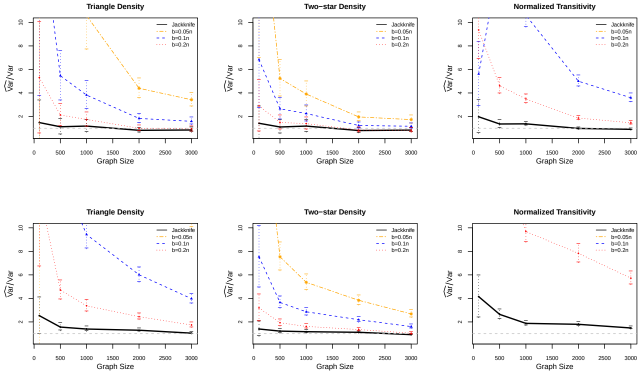
$$\text{Two star density} := \frac{\sum_{i,j<k,j,k\neq i} A_{ij}A_{ik}}{\binom{n}{3}\rho_n^2}$$



As a smooth function of count statistics, we use:

$$\text{Normalized transitivity} := \frac{\text{Triangle density}}{\text{Two star density}}$$

Simulation Study



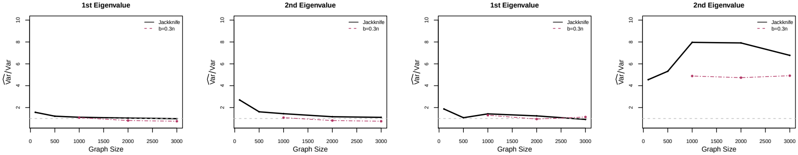
Ratio of VarJACK to true variance for triangle density, two-star density and transitivity in size= 100,500,1000,2000,3000 graphs simulated from the SBM (top) and GR2 (bottom), compared to subsampling with b= 0.05n,b= 0.1n,b= 0.2n variance estimation on the same graphs.

A note on Computation

- For each simulated network, we remove one node at a time, recalculate a statistic $Z_{n,i}$ on the graph with $(n-1)$ nodes left. Next we compute the jackknife estimate of the variance $\widehat{\text{Var}}_{\text{JACK}} := \sum_i (Z_{n,i} - \bar{Z}_n)^2$.
- It should be noted that for some statistics, jackknife can be implemented to reduce computation.
    - For example, in calculating triangles, we calculate the number of triangle on the whole graph once and the number of triangles each node is involved in from matrix manipulation.

Simulations Beyond Count Functional

Eigenvalues: Network Efron-Stein provides upper bound for any general statistics invariant to node-permutation.



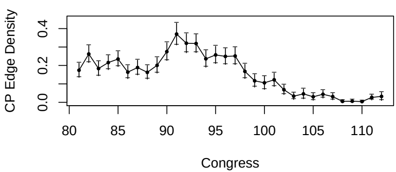
Ratio of Jackknife estimate $\widehat{\mathrm{Var}}_{\mathrm{JACK}}$ to true variance Var for first and second eigenvalues in size $n = 100, 500, 1000, 2000, 3000$ graphs simulated from SBM in (a) and (b) and GR(2) in (c) and (d), compared to subsampling with $b = 0.3n$.

Tentative evidence that our theory can be applied to statistics beyond count statistics, like eigenvalues. (Current work).

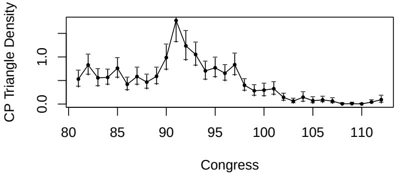
Network Jackknife: Real Data Application

1. Roll call vote data from the U.S. House of Representatives (vote view.com) from 1949 (commencement of the 81st Congress) to 2012 (adjournment of 112nd Congress).
2. Each Congress forms a network of representatives (nodes).
3. We have the number of agreements on bills (yay/yay or nay/nay)
4. For each network (Andris et al., 2015) compute a threshold, such that a randomly picked pair is more likely to be from the same party if their number of agreements is above it.
5. We build a unweighted graph using this threshold.
6. For each Congress, we calculate the normalized cross party edge density and cross party triangle density and construct CI from jackknife variance estimates.

Network Jackknife: Real Data Application



(A)



(B)

Cross party (A) edge density , and (B) triangle density and their CI's based on jackknife.

Takeaways for Jackknife

- For a broad class of network statistics, where we don't even have an asymptotic distribution, network jackknife provides an upper bound on the true variance, in average
- For subgraph counts and smooth functions of them, it is consistent.
- Easy to compute for counting small subgraphs.

Thank you! Any questions?

Andris, C., Lee, D., Hamilton, M. J., Martino, M., Gunning, C. E., and Selden, J. A. (2015). The rise of partisanship and super-cooperators in the u.s. house of representatives. *PLOS ONE*, 10(4):1–14.

Arvesen, J. N. (1969). Jackknifing U-statistics. *The Annals of Mathematical Statistics*, 40(6):2076–2100.

Bhattacharyya, S. and Bickel, P. J. (2015). Subsampling bootstrap of count features of networks. *Annals of Statistics*, 43:2384–2411.

Bickel, P. J., Chen, A., and Levina, E. (2011). The method of moments and degree distributions for network models. *Annals of Statistics*, 39:38–59.

Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., and Vesztergombi, K. (2008). Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801 – 1851.

Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., and Vesztergombi, K. (2012). Convergent sequences of dense graphs II: Multiway cuts and statistical physics. *Annals of Mathematics*, 176:151–219.

Crane, H. (2018). *Probabilistic Foundations of Statistical Network Analysis*. CRC Press, Boca Routon, FL.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(3):1–26.

Eldridge, J., Belkin, M., and Wang, Y. (2018). Unperturbed: spectral analysis beyond Davis-Kahan. In *Proceedings of Algorithmic Learning Theory (ALT)*, volume 83, pages 321–358.

Frank, O. and Snijders, T. (1994). Estimating the size of hidden populations using snowball sampling. *Journal of Official Statistics*, 10(1):53 – 67.

Füredi, Z. and Komlós, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica*, 1:233–241.

Gorochowski, T. E., Grierson, C. S., and di Bernardo, M. (2018). Organization of feed-forward loop motifs reveals architectural principles in natural and engineered networks. *Science Advances*, 4(3).

Green, A. and Shalizi, C. (2017). Bootstrapping Exchangeable Random Graphs. *arXiv e-prints*, page arXiv:1711.00813.

Heer, J. and Boyd, D. (2005). Vizster: Visualizing online social networks. In *Proceedings of the IEEE Conference on Information Visualization*.

Holland, P. W., Laskey, K., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, 5(2):109 – 137.

Koltchinskii, V. and Giné, E. (2001). Random matrix approximation of spectra of integral operators. *Bernoulli*, 6(1):1–24.

Lee, A. J. (1990). *U-statistics: Theory and Practice*. CRC Press, Boca Raton, Florida.

Levin, K. and Levina, E. (2019). Bootstrapping networks with latent space structure. *ArXiv e-prints*.

Lunde, R. and Sarkar, P. (2019). Subsampling sparse graphons under minimal assumptions. *ArXiv e-prints*.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics*, 22(4):2031–2050.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, pages 353–360.

Rhee, W. T. and Talagrand, M. (1986). Martingale inequalities and the jackknife estimate of variance. *Statistics & Probability Letters*, 4(1):5– 6.

Shao, J. and Tu, D. (1995). *The Jackknife and the Bootstrap*. Springer.

Tukey, J. W. (1958). Bias and confidence in not quite large samples (abstract). *The Annals of Mathematical Statistics*, 29:614.

Zhang, Y. and Xia, D. (2020). Edgeworth expansions for network moments. *arXiv preprint arXiv:2004.06615*.