A new way to think about citations

Alex Hayes

2020-11-17

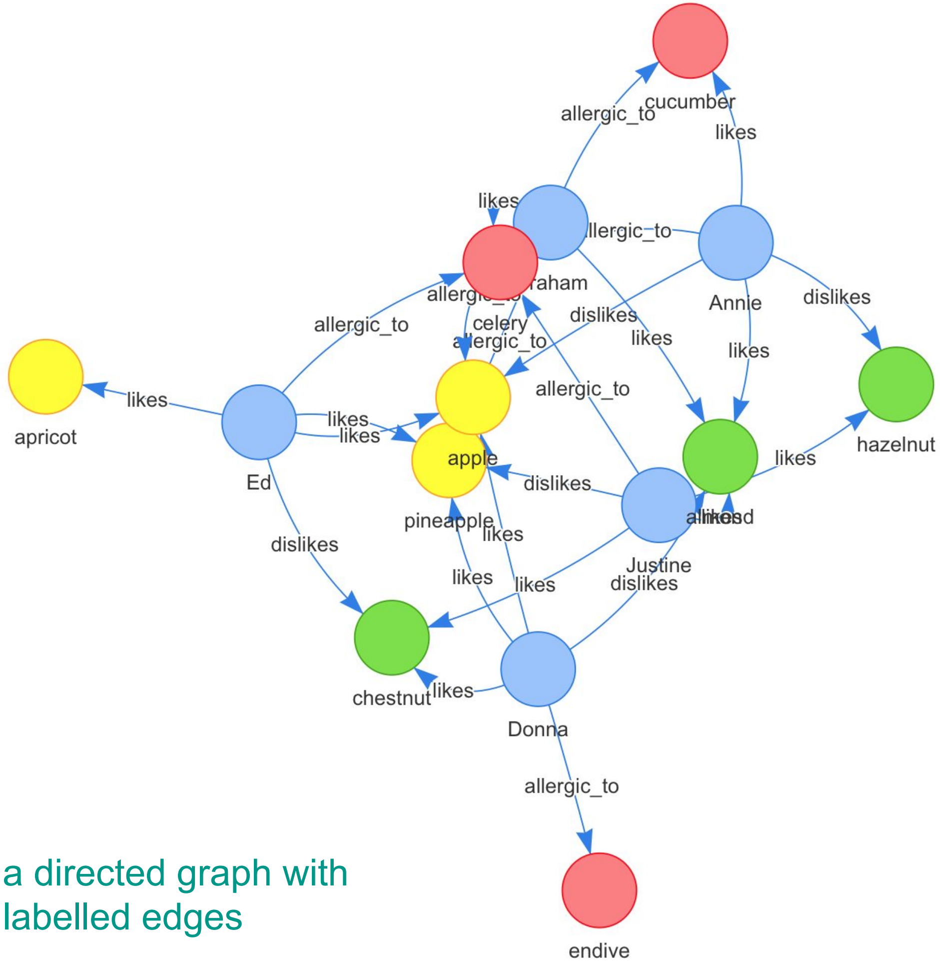
slides available @ https://bit.ly/citation-impute

Representing data as a graph

Graph = nodes + edges

Nodes are items under consideration

Edges are relationships between those items



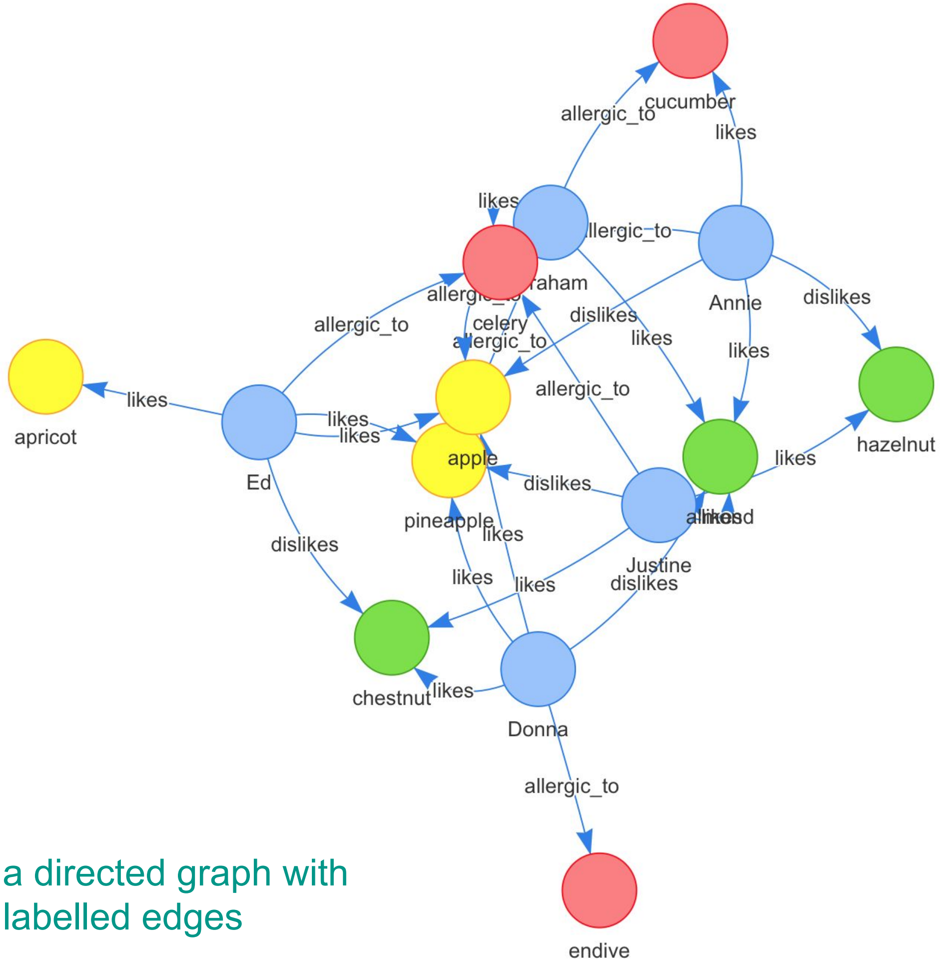
a directed graph with labelled edges

Graph = nodes + edges

Nodes are items under consideration

Edges are relationships between those items

We will think very carefully about this relationship



a directed graph with labelled edges

Goal: find clusters of documents that are topically similar

Goal: find clusters of documents that are topically similar

Edges should represent topical similarity between documents

Goal: find clusters of documents that are topically similar

Edges should represent topical similarity between documents

What data do we have about topical similarity??

How citations happen

Author writes a document

Document can cite previous related documents

Author publishes document

Published document is static and permanent

REFERENCES

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed Membership Stochastic Blockmodels. 34.

BOMMARITO II, M. J. and KATZ, D. M. (2009). Properties of the United States Code Citation Network. *arXiv:0911.1751 [physics]*.

CARMICHAEL, I., WUDEL, J., KIM, M. and JUSHCHUK, J. (2017). Examining the Evolution of Legal Precedent Through Citation Network Analysis. *NORTH CAROLINA LAW REVIEW* **96** 44.

CHAUDHURI, K., CHUNG, F. and TSIATAS, A. (2012). Spectral Clustering of Graphs with General Degrees in the Extended Planted Partition Model. 23.

CHEN, Y., JALALI, A., SANGHAVI, S. and XU, H. (2014). Clustering Partially Observed Graphs via Convex Optimization. *arXiv:1104.4803 [cs, stat]*.

CHO, J., KIM, D. and ROHE, K. (2018). Intelligent Initialization and Adaptive Thresholding for Iterative Matrix Completion; Some Statistical and Algorithmic Theory for Adaptive-Impute. *Journal of Computational and Graphical Statistics* 1–26.

CLOUGH, J. R., GOLLINGS, J., LOACH, T. V. and EVANS, T. S. (2015). Transitive Reduction of Citation Networks. *Journal of Complex Networks* **3** 189–203.

GERLACH, M., PEIXOTO, T. P. and ALTMANN, E. G. (2018). A Network Approach to Topic Models. *Science Advances* **4** eaaq1360.

HOERL, A. E. and KENNARD, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **12** 55–67.

JI, P. and JIN, J. (2014). Coauthorship and Citation Networks for Statisticians. 36.

LATOUCHE, P., BIRMELÉ, E. and AMBROISE, C. (2011). Overlapping Stochastic Block Models with Application to the French Political Blogosphere. *The Annals of Applied Statistics* **5** 309–336.

LEI, J. and RINALDO, A. (2015). Consistency of Spectral Clustering in Stochastic Block Models. *The Annals of Statistics* **43** 215–237.

MARTIN, T., BALL, B. and NEWMAN, M. E. J. (2016). Structural Inference for Uncertain Networks. *Physical Review E* **93** 012306.

MAZUMDER, R., HASTIE, T. and TIBSHIRANI, R. (2010). Spectral Regularization Algorithms for Learning Large Incomplete Matrices.

Citations contain information about topical similarity!

Traditional approach to clustering citation networks

Papers are nodes

Citations are directed edges

(Spectral) cluster

COAUTHORSHIP AND CITATION NETWORKS FOR STATISTICIANS

By Pengsheng Ji[†] and Jiashun Jin[‡]

University of Georgia[†] and Carnegie Mellon University[‡]

We have collected and cleaned two network data sets: Coauthorship and Citation networks for statisticians. The data sets are based on all research papers published in four of the top journals in statistics from 2003 to the first half of 2012. We analyze the data sets from many different perspectives, focusing on (a) productivity, patterns and trends, (b) centrality, and (c) community structures.

For (a), we find that over the 10-year period, both the average number of papers per author and the fraction of self citations have been decreasing, but the proportion of distant citations has been increasing. These findings are consistent with the belief that the statistics community has become increasingly more collaborative, competitive, and globalized.

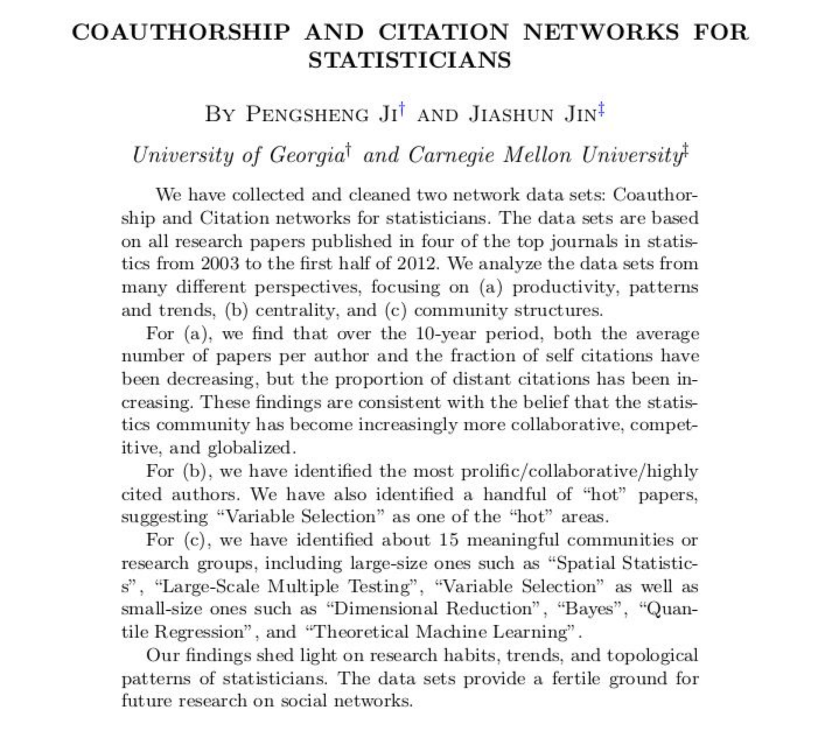
For (b), we have identified the most prolific/collaborative/highly cited authors. We have also identified a handful of "hot" papers, suggesting "Variable Selection" as one of the "hot" areas.

For (c), we have identified about 15 meaningful communities or research groups, including large-size ones such as "Spatial Statistics", "Large-Scale Multiple Testing", "Variable Selection" as well as small-size ones such as "Dimensional Reduction", "Bayes", "Quantile Regression", and "Theoretical Machine Learning".

Our findings shed light on research habits, trends, and topological patterns of statisticians. The data sets provide a fertile ground for future research on social networks.

Classic paper is Ji and Jin (2014)

Traditional approach to clustering citation networks



COAUTHORSHIP AND CITATION NETWORKS FOR STATISTICIANS

By Pengsheng Ji[†] and Jiashun Jin[‡]

University of Georgia[†] *and Carnegie Mellon University*[‡]

We have collected and cleaned two network data sets: Coauthorship and Citation networks for statisticians. The data sets are based on all research papers published in four of the top journals in statistics from 2003 to the first half of 2012. We analyze the data sets from many different perspectives, focusing on (a) productivity, patterns and trends, (b) centrality, and (c) community structures.

For (a), we find that over the 10-year period, both the average number of papers per author and the fraction of self citations have been decreasing, but the proportion of distant citations has been increasing. These findings are consistent with the belief that the statistics community has become increasingly more collaborative, competitive, and globalized.

For (b), we have identified the most prolific/collaborative/highly cited authors. We have also identified a handful of "hot" papers, suggesting "Variable Selection" as one of the "hot" areas.

For (c), we have identified about 15 meaningful communities or research groups, including large-size ones such as "Spatial Statistics", "Large-Scale Multiple Testing", "Variable Selection" as well as small-size ones such as "Dimensional Reduction", "Bayes", "Quantile Regression", and "Theoretical Machine Learning".

Our findings shed light on research habits, trends, and topological patterns of statisticians. The data sets provide a fertile ground for future research on social networks.

Papers are nodes

Citations are directed edges

(Spectral) cluster

We think we can do better here

Classic paper is Ji and Jin (2014)

Older documents can't cite newer documents!

TECHNOMETRICS VOL. 12, NO. 1 FEBRUARY 1970

Ridge Regression: Biased Estimation for Nonorthogonal Problems

ARTHUR E. HOERL AND ROBERT W. KENNARD
University of Delaware and E. I. du Pont de Nemours & Co.

In multiple regression it is shown that parameter estimates based on minimum residual sum of squares have a high probability of being unsatisfactory, if not incorrect, if the prediction vectors are not orthogonal. Proposed is an estimation procedure based on adding small positive quantities to the diagonal of $\mathbf{X}'\mathbf{X}$. Introduced is the ridge trace, a method for showing in two dimensions the effects of nonorthogonality. It is then shown how to augment $\mathbf{X}'\mathbf{X}$ to obtain biased estimates with smaller mean square error.

0. INTRODUCTION

Consider the standard model for multiple linear regression, $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where it is assumed that $\mathbf{X}$ is $(n \times p)$ and of rank p, $\boldsymbol{\beta}$ is $(p \times 1)$ and unknown, $E[\boldsymbol{\varepsilon}] = \mathbf{0}$, and $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \sigma^2\mathbf{I}_n$. If an observation on the factors is denoted by $\mathbf{x}_v = \{x_{1v}, x_{2v}, \ldots, x_{pv}\}$, the general form $\mathbf{X}\boldsymbol{\beta}$ is $\{\sum_{i=1}^{p} \beta_i\theta_i(\mathbf{x}_v)\}$ where the θ_i are functions free of unknown parameters.

The usual estimation procedure for the unknown $\boldsymbol{\beta}$ is Gauss-Markov—linear functions of $\mathbf{Y} = \{y_v\}$ that are unbiased and have minimum variance. This estimation procedure is a good one if $\mathbf{X}'\mathbf{X}$, when in the form of a correlation matrix, is nearly a unit matrix. However, if $\mathbf{X}'\mathbf{X}$ is not nearly a unit matrix, the least squares estimates are sensitive to a number of "errors." The results of these errors are critical when the specification is that $\mathbf{X}\boldsymbol{\beta}$ is a true model. Then the least squares estimates often do not make sense when put into the context of the physics, chemistry, and engineering of the process which is generating the data. In such cases, one is forced to treat the estimated predicting function as a black box or to drop factors to destroy the correlation bonds among the $\mathbf{X}_i$ used to form $\mathbf{X}'\mathbf{X}$. Both these alternatives are unsatisfactory if the original intent was to use the estimated predictor for control and optimization. If one treats the result as a black box, he must caution the user of the model not to take partial derivatives (a useless caution in practice), and in the other case

Ridge paper (1970) could not cite LASSO paper (1996)

Ridge paper is similar to LASSO paper

J. R. Statist. Soc. B (1996)
58, *No.* 1, *pp.* 267–288

Regression Shrinkage and Selection via the Lasso

By ROBERT TIBSHIRANI†
University of Toronto, Canada

[Received January 1994. Revised January 1995]

SUMMARY

We propose a new method for estimation in linear models. The 'lasso' minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. Because of the nature of this constraint it tends to produce some coefficients that are exactly 0 and hence gives interpretable models. Our simulation studies suggest that the lasso enjoys some of the favourable properties of both subset selection and ridge regression. It produces interpretable models like subset selection and exhibits the stability of ridge regression. There is also an interesting relationship with recent work in adaptive function estimation by Donoho and Johnstone. The lasso idea is quite general and can be applied in a variety of statistical models: extensions to generalized regression models and tree-based models are briefly described.

Keywords: QUADRATIC PROGRAMMING; REGRESSION; SHRINKAGE; SUBSET SELECTION

1. INTRODUCTION

Consider the usual regression situation: we have data $(\mathbf{x}^i, y_i)$, $i = 1, 2, \ldots, N$, where $\mathbf{x}^i = (x_{i1}, \ldots, x_{ip})^T$ and y_i are the regressors and response for the ith observation. The ordinary least squares (OLS) estimates are obtained by minimizing the residual squared error. There are two reasons why the data analyst is often not satisfied with the OLS estimates. The first is *prediction accuracy*: the OLS estimates often have low bias but large variance; prediction accuracy can sometimes be improved by shrinking or setting to 0 some coefficients. By doing so we sacrifice a little bias to reduce the variance of the predicted values and hence may improve the overall prediction accuracy. The second reason is *interpretation*. With a large number of predictors, we often would like to determine a smaller subset that exhibits the strongest effects.

The two standard techniques for improving the OLS estimates, subset selection and ridge regression, both have drawbacks. Subset selection provides interpretable models but can be extremely variable because it is a discrete process—regressors are either retained or dropped from the model. Small changes in the data can result in very different models being selected and this can reduce its prediction accuracy. Ridge regression is a continuous process that shrinks coefficients and hence is more stable: however, it does not set any coefficients to 0 and hence does not give an easily interpretable model.

We propose a new technique, called the *lasso*, for 'least absolute shrinkage and selection operator'. It shrinks some coefficients and sets others to 0, and hence tries to retain the good features of both subset selection and ridge regression.

Wacky idea

We know: physical presence of citations

Want to know: similarities between documents

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ ? & 0 & 0 & 1 & 0 & 0 \\ ? & ? & 0 & 0 & 0 & 1 \\ ? & ? & ? & 0 & 1 & 0 \\ ? & ? & ? & ? & 0 & 1 \\ ? & ? & ? & ? & ? & 0 \end{bmatrix}$$

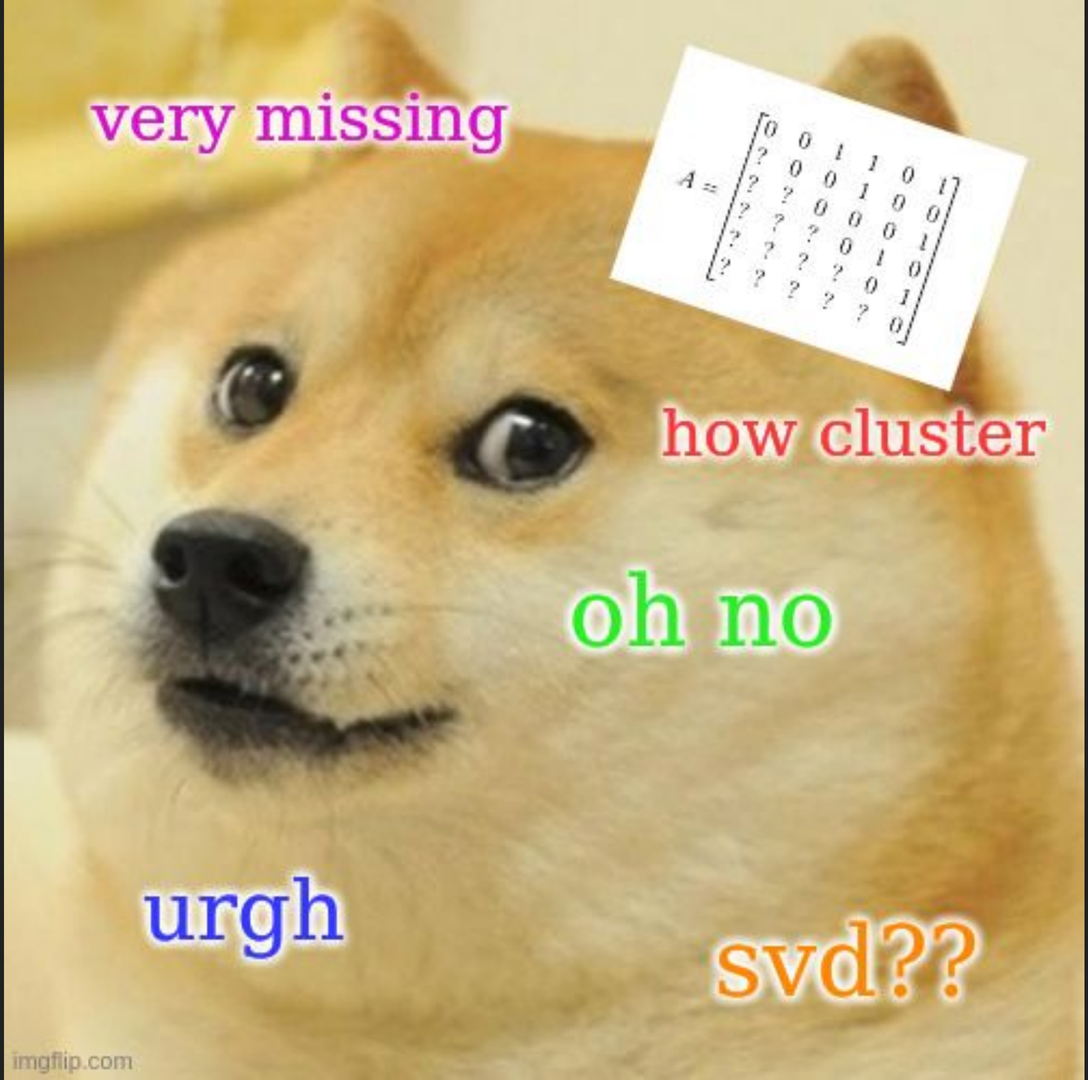
Use partially observed adjacency matrix of “similarity network”

All the options

$$A = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ ? & 0 & 0 & 1 & 0 & 0 \\ ? & ? & 0 & 0 & 0 & 1 \\ ? & ? & ? & 0 & 1 & 0 \\ ? & ? & ? & ? & 0 & 1 \\ ? & ? & ? & ? & ? & 0 \end{bmatrix}}_{\text{directed similarity relationship}}, \quad A = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\text{physical citations}}, \quad A = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}}_{\text{undirected similarities}}$$

* Our idea only makes sense if you think citations should be directed edges (they should).

An estimator


very missing
$$A = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ ? & 0 & 0 & 1 & 0 & 0 \\ ? & ? & 0 & 0 & 0 & 1 \\ ? & ? & ? & 0 & 1 & 0 \\ ? & ? & ? & ? & 0 & 1 \\ ? & ? & ? & ? & ? & 0 \end{bmatrix}$$
how cluster
oh no
urgh
svd??
imgflip.com

Spectral clustering

1. Take SVD of A ~ U D V’
2. Clusters rows of A and rows of V’
3. Profit

Spectral clustering

SVD doesn't work
with missing data!

1. Take ~~SVD~~ of A ~ U D V'
2. Clusters rows of A and rows of V'
3. Profit

Spectral clustering

Works with missing data!

1. Use matrix completion on A ~ U D V’
2. Clusters rows of A and rows of V’
3. Profit

By matrix completion I mean `AdaptiveImpute`

Consistent estimator of U, V under some assumptions

A lot like `SoftALS`

O(observed edges * rank) runtime

Algorithm 1: `AdapativeImpute`

Input: M, y, r and $\varepsilon > 0$

$Z^{(1)} \leftarrow$ `AdaptiveInitialize`(M, y, r)

repeat

$\tilde{M}^{(t)} \leftarrow P_\Omega(M) + P_\Omega^\perp(Z_t)$

$\hat{V}_i^{(t)} \leftarrow \mathbf{v}_i(\tilde{M}^{(t)})$ for $i = 1, ..., r$

$\hat{U}_i^{(t)} \leftarrow \mathbf{u}_i(\tilde{M}^{(t)})$ for $i = 1, ..., r$

$\tilde{\alpha}^{(t)} \leftarrow \frac{1}{d-r} \sum_{i=r+1}^{d} \boldsymbol{\lambda}_i^2(\tilde{M}^{(t)})$

$\hat{\lambda}_i^{(t)} \leftarrow \sqrt{\boldsymbol{\lambda}_i^2(\tilde{M}^{(t)}) - \tilde{\alpha}^{(t)}}$ for $i = 1, ..., r$

$Z^{(t+1)} \leftarrow \sum_{i=1}^{r} \hat{\lambda}_i^{(t)} \hat{U}_i^{(t)} \hat{V}_i^{(t)^T}$

$t \leftarrow t + 1$

until $\|Z_{t+1} - Z_t\|_F^2 / \|Z_{t+1}\|_F$

return $\hat{\lambda}_i^{(t)}, \hat{U}_i^{(t)}, \hat{V}_i^{(t)}$ *for* $i = 1, ..., r$

By matrix completion I mean `AdaptiveImpute`

Consistent estimator of U, V under some assumptions

A lot like `SoftALS`

O(observed edges * rank) runtime

Algorithm 1: `AdapativeImpute`

Input: M, y, r and $\varepsilon > 0$

$Z^{(1)} \leftarrow$ `AdaptiveInitialize`(M, y, r)

repeat

$\quad \tilde{M}^{(t)} \leftarrow P_\Omega(M) + P_\Omega^\perp(Z_t)$

$\quad \hat{V}_i^{(t)} \leftarrow \mathbf{v}_i(\tilde{M}^{(t)})$ for $i = 1, ..., r$

$\quad \hat{U}_i^{(t)} \leftarrow \mathbf{u}_i(\tilde{M}^{(t)})$ for $i = 1, ..., r$

$\quad \tilde{\alpha}^{(t)} \leftarrow \frac{1}{d-r} \sum_{i=r+1}^{d} \boldsymbol{\lambda}_i^2(\tilde{M}^{(t)})$

$\quad \hat{\lambda}_i^{(t)} \leftarrow \sqrt{\boldsymbol{\lambda}_i^2(\tilde{M}^{(t)}) - \tilde{\alpha}^{(t)}}$ for $i = 1, ..., r$

$\quad Z^{(t+1)} \leftarrow \sum_{i=1}^{r} \hat{\lambda}_i^{(t)} \hat{U}_i^{(t)} \hat{V}_i^{(t)^T}$

$\quad t \leftarrow t + 1$

until $||Z_{t+1} - Z_t||_F^2 / ||Z_{t+1}||_F$

return $\hat{\lambda}_i^{(t)}, \hat{U}_i^{(t)}, \hat{V}_i^{(t)}$ *for* $i = 1, ..., r$

By matrix completion I mean `AdaptiveImpute`

Consistent estimator of U, V under some assumptions

A lot like `SoftALS`

O(observed edges * rank) runtime

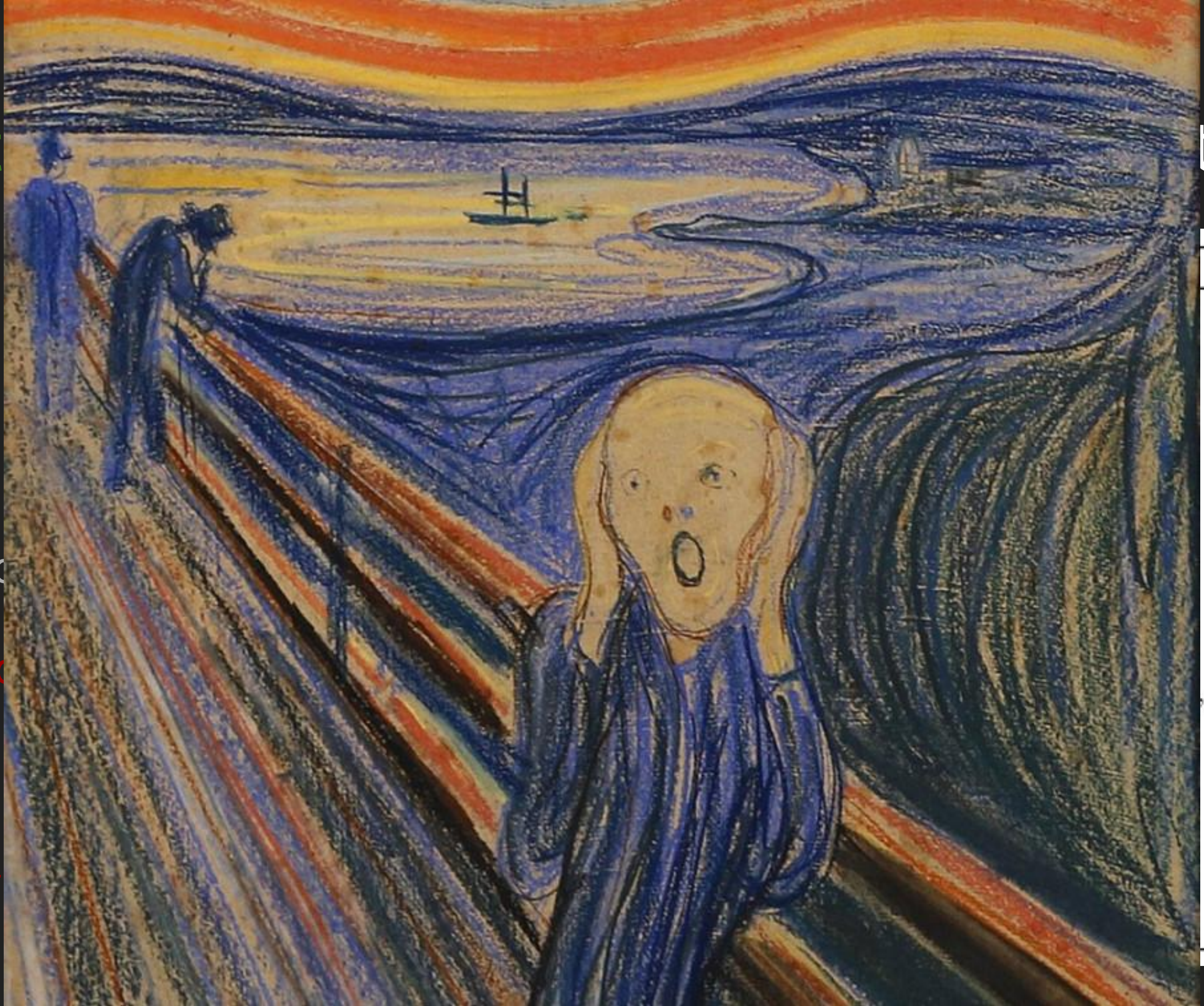
Fully half of all edges are observed!
Not sparse!

Algorithm 1: `AdapativeImpute`

Input: M, y, r and $\varepsilon > 0$

$Z^{(1)} \leftarrow$ `AdaptiveInitialize`(M, y, r)
repeat
  $\tilde{M}^{(t)} \leftarrow P_\Omega(M) + P_\Omega^\perp(Z_t)$
  $\hat{V}_i^{(t)} \leftarrow \mathbf{v}_i(\tilde{M}^{(t)})$ for $i = 1, ..., r$
  $\hat{U}_i^{(t)} \leftarrow \mathbf{u}_i(\tilde{M}^{(t)})$ for $i = 1, ..., r$
  $\tilde{\alpha}^{(t)} \leftarrow \frac{1}{d-r} \sum_{i=r+1}^{d} \boldsymbol{\lambda}_i^2(\tilde{M}^{(t)})$
  $\hat{\lambda}_i^{(t)} \leftarrow \sqrt{\boldsymbol{\lambda}_i^2(\tilde{M}^{(t)}) - \tilde{\alpha}^{(t)}}$ for $i = 1, ..., r$
  $Z^{(t+1)} \leftarrow \sum_{i=1}^{r} \hat{\lambda}_i^{(t)} \hat{U}_i^{(t)} \hat{V}_i^{(t)^T}$
  $t \leftarrow t + 1$
until $\|Z_{t+1} - Z_t\|_F^2 / \|Z_{t+1}\|_F$
return $\hat{\lambda}_i^{(t)}, \hat{U}_i^{(t)}, \hat{V}_i^{(t)}$ for $i = 1, ..., r$

By mat te



Consistent
some assu

A lot like Sc

O(observed

Fully half of
Not sparse!

Computational complexity

Computational complexity

We got it down to $O(\text{number citations} + \text{nodes} * \text{rank}^2)$ runtime

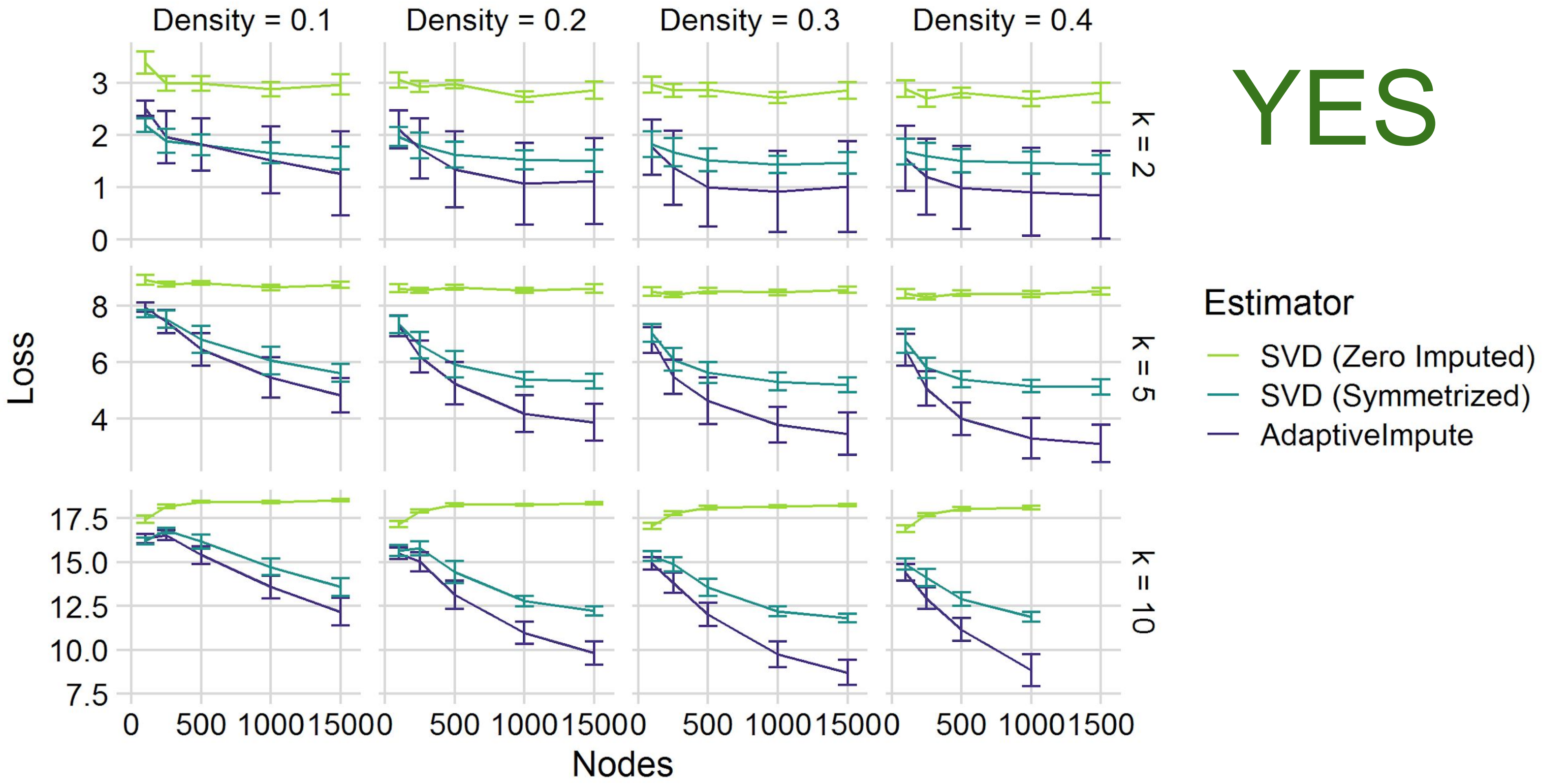
Computational complexity

We got it down to $O(\text{number citations} + \text{nodes} * \text{rank}^2)$ runtime

The details aren't very exciting

Does it work?

AdaptiveImpute recovers principal subspaces better than SVD



YES

Simulation details

Procedure:

Draw from stochastic blockmodel

Throw out lower triangle

Check how well we recover U and V’

Compare to:

SVD on symmetrized A

SVD on “physical similarity” A

$$A = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ ? & 0 & 0 & 1 & 0 & 0 \\ ? & ? & 0 & 0 & 0 & 1 \\ ? & ? & ? & 0 & 1 & 0 \\ ? & ? & ? & ? & 0 & 1 \\ ? & ? & ? & ? & ? & 0 \end{bmatrix}}_{\text{directed similarity relationship}}, \quad A = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\text{physical citations}}, \quad A = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}}_{\text{undirected similarities}}$$

Navel gazing

Web of Science citation data

281 million papers from 1900 to 2019

consider 125 journals on probability and statistics

leaves ~200k papers with 1.5 million citations

Methods

form the partially observed adjacency matrix A

use `AdaptiveImpute` to estimate U and V (rank 30)

varimax rotate U and V

interpret with bff

Comprehensive breakdown of statistics literature

factor	words
y1	supersaturated, designs, optimal, martingale, ruin, transport, hedging, construction, ranked, options
y10	fuzzy, numbers, trapezoidal, designs, approximations, optimal, approximation, preserving, triangular, experiments
y11	depth, multivariate, robust, skew, functional, breakdown, outlier, location, high, principal
y12	trials, clinical, sequential, adaptive, size, group, stage, interim, treatment, phase
y13	high, false, discovery, multiple, dimensional, testing, causal, sparse, covariance, propensity
y14	coherent, systems, residual, components, order, comparisons, distributions, lifetimes, out, ordering
y15	spatial, temporal, spatio, covariance, bayesian, fields, space, disease, gaussian, datasets
y16	treatment, regimes, causal, learning, propensity, optimal, individualized, randomized, trials, dynamic
y17	meta, longitudinal, analysis, models, effects, regression, mixed, trials, data, missing
y18	regression, selection, dimensional, high, screening, variable, lasso, quantile, ultrahigh, sparse
y19	designs, split, optimal, plot, experiments, factorial, design, d, fractional, surface
y2	volatility, processes, estimation, density, realized, jumps, frequency, nonparametric, noise, levy
y20	functional, regression, nonparametric, data, linear, smoothing, kernel, density, spline, estimation
y21	resolution, calibration, multivariate, curve, pls, way, analysis, parafac, least, squares
y22	series, extreme, dependence, autoregressive, copula, copulas, tail, supersaturated, processes, extremes
y23	designs, trials, clinical, adaptive, urn, randomization, optimal, design, sequential, coin
y24	dimension, reduction, sliced, sufficient, inverse, regression, index, propensity, single, causal
y25	heterogeneous, order, ordering, comparisons, systems, statistics, parallel, components, from, variables
y26	designs, level, factorial, aberration, supersaturated, minimum, fractional, lower, discrepancy, 2
y27	ranked, set, sampling, judgment, samples, distribution, order, post, censored, statistics
y28	dose, phase, i, finding, trials, clinical, continual, reassessment, design, bayesian
y29	control, charts, monitoring, chart, cusum, ewma, process, profiles, exponentially, change
y3	reinsurance, investment, ruin, risk, optimal, levy, insurers, insurer, dividend, spectrally
y30	group, testing, proportions, confidence, density, intervals, nonparametric, recapture, deconvolution, estimation
y4	fuzzy, coalescent, trees, branching, brownian, planar, coalescents, random, maps, scaling

factor	words
z1	supersaturated, designs, optimal, martingale, ruin, transport, hedging, construction, ranked, options
z10	fuzzy, numbers, trapezoidal, designs, approximations, optimal, approximation, preserving, triangular, experiments
z11	depth, multivariate, robust, skew, functional, breakdown, outlier, location, high, principal
z12	trials, clinical, sequential, adaptive, size, group, stage, interim, treatment, phase
z13	high, false, discovery, multiple, dimensional, testing, causal, sparse, covariance, propensity
z14	coherent, systems, residual, components, order, comparisons, distributions, lifetimes, out, ordering
z15	spatial, temporal, spatio, covariance, bayesian, fields, space, disease, gaussian, datasets
z16	treatment, regimes, causal, learning, propensity, optimal, individualized, randomized, trials, dynamic
z17	meta, longitudinal, analysis, models, effects, regression, mixed, trials, data, missing
z18	regression, selection, dimensional, high, screening, variable, lasso, quantile, ultrahigh, sparse
z19	designs, split, optimal, plot, experiments, factorial, design, d, fractional, surface
z2	volatility, processes, estimation, density, realized, jumps, frequency, nonparametric, noise, levy
z20	functional, regression, nonparametric, data, linear, smoothing, kernel, density, spline, estimation
z21	resolution, calibration, multivariate, curve, pls, way, analysis, parafac, least, squares
z22	series, extreme, dependence, autoregressive, copula, copulas, tail, supersaturated, processes, extremes
z23	designs, trials, clinical, adaptive, urn, randomization, optimal, design, sequential, coin
z24	dimension, reduction, sliced, sufficient, inverse, regression, index, propensity, single, causal
z25	heterogeneous, order, ordering, comparisons, systems, statistics, parallel, components, from, variables
z26	designs, level, factorial, aberration, supersaturated, minimum, fractional, lower, discrepancy, 2
z27	ranked, set, sampling, judgment, samples, distribution, order, post, censored, statistics
z28	dose, phase, i, finding, trials, clinical, continual, reassessment, design, bayesian
z29	control, charts, monitoring, chart, cusum, ewma, process, profiles, exponentially, change
z3	reinsurance, investment, ruin, risk, optimal, levy, insurers, insurer, dividend, spectrally
z30	group, testing, proportions, confidence, density, intervals, nonparametric, recapture, deconvolution, estimation
z4	fuzzy, coalescent, trees, branching, brownian, planar, coalescents, random, maps, scaling

Thank you! Questions?

Stay in touch

https://www.alexpghayes.com/

https://twitter.com/alexpghayes

Slides available at

https://bit.ly/citation-impute

Code available at

https://github.com/RoheLab/fastadi

https://github.com/RoheLab/vsp