Final Report

Rachel Springer (rs4127)

December 21, 2023

Abstract:

Despite a rise in the visibility of labor movements in recent years, little remains known quantitatively about the dynamics of workplace organizing. This study aims to understand what worker, worksite, and organizing features were correlated to an individual signing a membership card in the 2010-2015 OUR Walmart organizing campaign. Drawing on social movement theory, we understand both that networks play an essential role in movement participation and that these dynamics can be different across different social contexts. Previous work on analyzing the dynamics of workplace organizing used a pooled OLS regression model to analyze the importance of network-driven organizing. Here we use an augmented version of this dataset through a Bayesian hierarchical model to account for both intra- and inter-workplace variation in organizing effects, seeking to develop strategies and recommendations to assess organizing strategies for labor organizations. A hierarchical Bayesian logistic regression model was employed to identify correlations between workplace and worker attributes and a workers choice to sign a membership card for the organization while allowing for variation across workplaces. A Gaussian mixture model was then run on the resulting workplace coefficients to identify potential clusters in these predictors. The results show that an individuals choice to vote was most dependent on workplace factors, with network-driven organizing, number of organizer conversations, average workplace centrality correlating positively to an individual choosing to sign a card, while mid-length campaigns, large worksites, and workplaces located in zip codes with a higher Latino population were correlated negatively with an individuals choice to sign a card. Furthermore, a clustering of worksite-specific coefficients found that in larger worksites organizer attention to the individual worker was more important in predicting their choice to sign a card. These results suggest an importance of group and social factors over individual features, even when measuring on an individual-level outcome. Based on these results, we suggest a focus on group-level interventions on workplace organizing with consciousness of differing dynamics in different workplace scales and a systems theory approach to understanding workplace organizing.

Methods:

Data was obtained from an OUR Walmart organizing campaign from 2010-2015, containing 18 worksite-level features and 7 worker level features, encompassing 120 worksites and 16,553 workers. Worksite-level features were Lasso-selected [Sheperd et al 2023], including store-level demographics and features of workplace networks, which were reconstructed from organizer notes. The workplace features included in the regression where percent Black in workplace ZIP, percent Latino in workplace ZIP, percent male at workplace, mean AGI in ZIP, network-driven organizing (operationalized as the rank-correlation between organizer notes and worker centrality), mobilizing (operationalized as the correlation between the number of organizer conversations and how early in the campaign the worker was first contacted), campaign length, number of organizer conversations, number workers contacted, number of workers discovered, number of edges in the workplace network, mean degree of the workplace network, centrality variance in the workplace network, and mean centrality in the workplace network. The features were log transformed and regularized

before performing inference.

A secondary dataset was also constructed to collect worker-level features. For each worker, data was collected from organizer notes and workplace data including gender, the quartile in which the worker had their first conversation with an organizer, their eigenvector centrality in the workplace network, and the number of organizer notes that referenced them (an operationalization of worker effort).

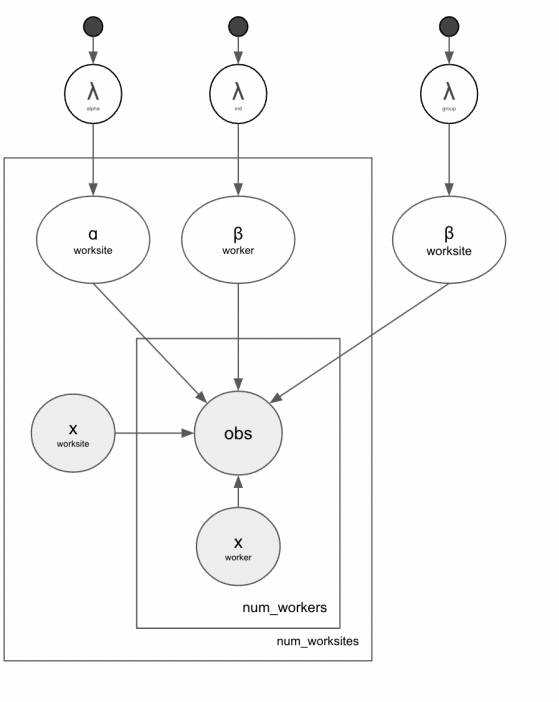


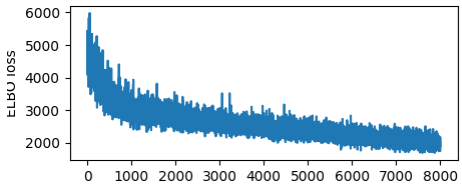
Fig 1

A Bayesian hierarchical logistic regression model was used to represent the data (Fig 1) due to its ability to capture variability at both an individual level and group level, and in more accurately representing the structure of the data [Gelman and Hill, 2007]. The model predicts whether a worker signed a membership card as a binary outcome drawn from a Bernoulli distribution. The logits come from a linear combination the worker's of individual and worksite features and the worker- and worksite- level coefficients that correspond that worksite. Coefficients for worker-level data were drawn for each worksite from a zero-mean normal distribution with a shared scale parameter, while the coefficients for worksite-level data were shared across all worksites.The model was set up in Pyro and optimized through stochastic variational inference using a mean-field parameterization.

In setting up the model, priors were chosen to be non-informative and flexible, representing the lack of prior information on the scale for the coefficients. A Half-Cauchy distribution with a scale of 5 was used to parameterize the distributions for the variance. These were used as the scale parameter in zero-mean normal distributions, centered for interpretability.

Various experiments were run to determine the parameters for stochastic variational inference, first on a subsample of the dataset and then the full dataset. The learning rate proved to be particularly

relevant: low learning rates (of 0.001 and below) resulted in frequent fluctuations in the ELBO loss as it got stuck in local minima. High learning rates avoided this problem and bypassed these local minima, but got stuck as they approached the solution, fluctuating at a plateau. Learning rates around 0.01 performed well, but still resulted in some fluctuating as the ELBO plateaued. To mitigate this, a decaying learning rate was implemented. Initially a decaying rate was attempted with ClippedAdam, but the clipped gradients resulted in upwards curve to ELBO loss. Running for 80,000 steps for completeness with an initial learning rate of 0.015 and a decay factor of 0.9995 (or 0.95 per 100 steps) led to convergence at 18,000 steps, which was used in subsequent analyses.



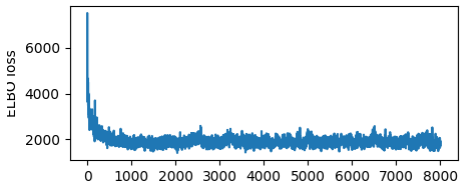
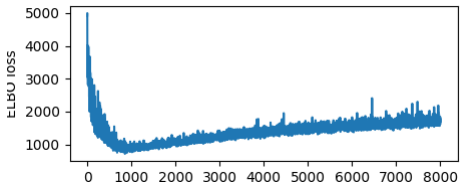


Fig 2 Low lr, Fig 3 High lr



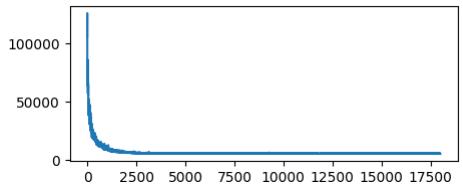


Fig 4 ClippedAdam, Fig 5 Convergence with Decaying LR

A posterior predictive check conducted on held out data suggested a good fit for the model, acheiving an accuracy score of 0.9356 (compared to a baseline of 0.8274 when predicting only 0) and an AUC of 0.7688

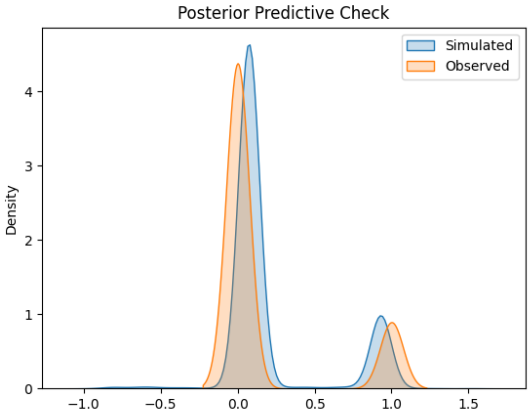


Fig 5, Kernel Density Estimation Plots on Posterior Predictive

After running the regression, a Gaussian Mixture Model was used to identify potential clusters in regression parameters across workplaces. A Gibbs sampler was used to identify the means and assignments of each workplace-coefficient pair. The sampler was run on the parameters σ = 0.0001, α_0= 1, η= 0.1, K= 2, with the number of iterations set to 400. This number of iterations led to convergence of the log joint. σ was set through experimentation- only smaller sigmas were able to effectively converge to cluster the data. The log likelihood of the joint plateaud after 2 clusters, so k=2 was used to analyze the data.

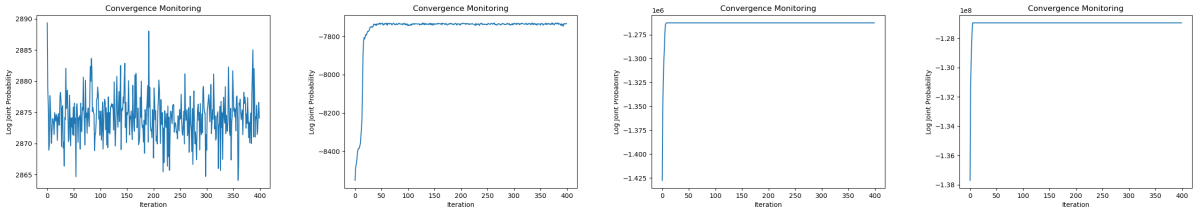


Fig 6, Convergence of Log Joint after 400 Iterations, σ = 0.01, 0.001, 0.0001, 0.00001

Regression Results:

Shared Coefficient	Value
β_{worker} scale	0.006294
β_{worksite} scale	2.141032
α mean	-1.831840
Percent Black in ZIP	0.293133
Percent Latino in ZIP	-1.890090
Percent male	-0.598283
Log(Mean AGI in ZIP)	-0.462811
NDO (Eigen., Pearson)	2.782533
Mobilizing (Pearson)	0.907776
Campaign length (Quint. 1)	0.221247
Campaign length (Quint. 2)	-0.047250
Campaign length (Quint. 3)	-1.972645
Campaign length (Quint. 4)	-0.261196
Campaign length (Quint. 5)	0.847465
Log(No. organizer convers.)	1.604383
Log(No. workers contacted)	-0.060864
Log(No. workers discovered)	-1.711616
Log(No. edges)	-0.215173
Log(Mean degree)	-0.783317
Log(Centrality variance)	0.918744
Log(Centrality mean)	1.153290

Worksite-specific Coefficients	median	mean	std
Is Male	-0.001070	-0.000121	0.005464
Time To First Convo (Quart. 1)	0.000468	0.000673	0.005359
Time To First Convo (Quart. 2)	-0.000400	-0.000152	0.005429
Time To First Convo (Quart. 3)	-0.000638	-0.000188	0.005675
Time To First Convo (Quart. 4)	0.001248	0.000292	0.005494
Log(Eigen. Centrality)	-0.000908	-0.000569	0.007155
No. Notes	0.000152	0.000353	0.007465

Of the shared coefficients for worksite-level features, network-driven organizing (NDO) acted as the strongest predictor for an individual's choice to sign a membership card, supporting previous research on the topic that it is an effective organizing tactic. The number of organizer conversations also positively correlates with a workers choice to sign a membership card, likely indicating the importance of such conversations to getting workers to sign. Workplace networks with a high average centrality were also correlated with an individual signing a card, perhaps indicating that knowing ones coworkers makes one more likely to be willing to sign a membership card, or because well-documented networks correlated with network-driven organizing or organizer effort, which also seem to predict outcomes. The positive correlation between workplace mobilizing and a card being signed and centrality variance and a card being signed may also be for the same reason.

On the other hand, we see stronger negative correlations with workplaces that have a high percentage

of Latino people in the ZIP code, have mid-length (quintile 3) campaigns, and have a high number of workers discovered. This may be explained by different cultural and material experiences with labor organizations across cultures. Mid-length campaigns may have neither the obvious support of campaigns that are shorter nor the increased effort of longer campaigns, and perhaps less attention is given to individuals or it is harder to organize in larger workplaces.

The small scale for β_{worksite} implies little variance between workplaces and smaller coefficients (mostly clustered around 0). These results suggest that worksite level features a better predictor of card signing than individual attributes, and that this choice, even measured at an individual outcome, is heavily reliant on social factors.

Regression Coefficient Clustering Results:

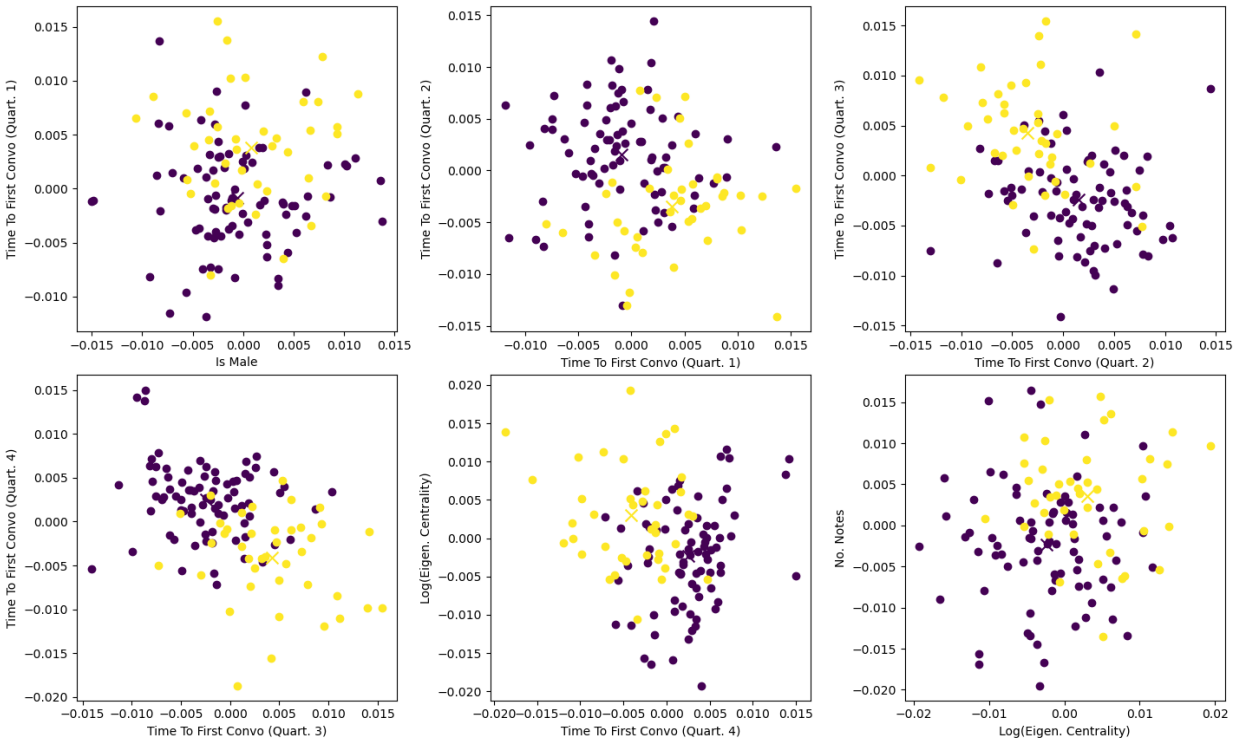


Fig 6, Regression Coeffficient Clusters

Plotting these cluster assignments against workplace features below, we find that while most features' values are evenly distributed across the clusters, there is some stratification in the number of workers contacted and the number of workers discovered in the cluster assignments. Namely, worksites with these features that correspond to larger worksites had a more positive correlation between the number of organizer notes on the worker and card signing, and earlier contact and card signing. This could indicate the the dynamics of larger worksite mean that organizer attention is more important or perhaps more rare in such environments.

Fig 7, Worksite Feature By Cluster

Conclusion

Findings suggest worksite level features are the most important predictors of an individuals voting outcome. Network-driven organizing, number of organizer conversations, high average centrality strongest predictors of an individuals choice to sign a card, while mid-length campaigns, large worksites, and workplaces located in zip codes with a higher Latino population were negatively correlated with an individuals choice to vote. Furthermore, worksite size affects dynamics, with larger worksites featuring a stronger correlation between the number of organizer notes on the worker and an earlier contact times- both operationalizations of organizer attention to the worker. Further research is suggested to look into how worksite organizing dynamics change across different worksite sizes. These results suggest the effectiveness and importance of organizer tactics over inherent workplace features, and the importance of social factors over individual features, even when measuring on an individual-level outcome. This work, however, is limited to correlational relationships, and may have missed a salient or unmeasured characteristic that determines outcomes. Further work with additional measurements, and working outside of a mean-field parameterization to better capture the relationships and dependencies between variables could serve to add on to these results.