

INTRODUCTION

- A conditional independence graph is a concise representation of pairwise conditional independence among many variables.
- We present a general framework for estimating pairwise conditional independence relationships among variables.

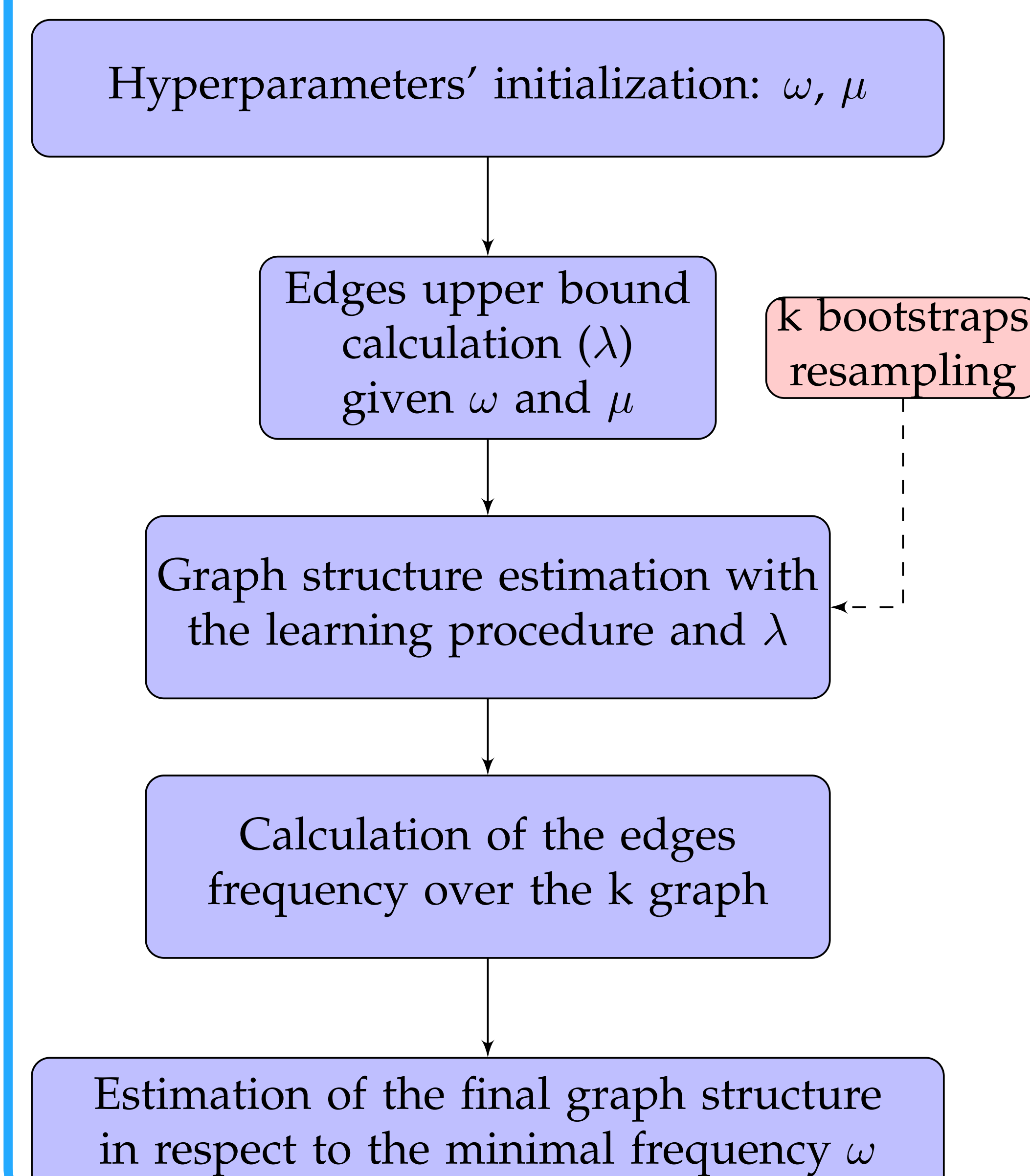
STABILITY SELECTION ALGORITHM

- Number of edges is a tuning parameter in any graphical model estimator.
- No obvious number constitutes a good choice.
- Stability Selection helps choosing the parameter (λ) with respect to a bound on the expected number of false positives μ :

$$\lambda = \lfloor \sqrt{(2\omega - 1) \cdot \mu \cdot p \cdot (p - 1)/2} \rfloor$$

where ω is a threshold of the minimum relative frequency of edges and p is the number of variables.

Pseudo code



ALGORITHMS

Graphical Random Forest (GRaFo)

Edges ranking scheme based on Random Forests' permutation importance measure.

Importance measure :

Predictor's relevance based on error difference between a regular Random Forest fit and a Random Forest fit within which one predictor's values are permuted.

Main steps :

1. Perform RF regression of a single variable on the remaining $p - 1$ variables.
2. Calculate predictors' importance measure.
3. Repeat steps 1 and 2 for every p variables.
4. Define the (i, j) pair as i the outcome and j the predictor.
5. Best λ ranking pairs are added as edges.

Graphical Lasso (GLasso)

Estimation of the regularized precision matrix Λ under multivariate normal hypothesis:

$$\hat{\Lambda}_{GL} = \operatorname{argmax}_{\Lambda} \{ \log |\Lambda| - \operatorname{tr}(S\Lambda) - \rho \|\Lambda\|_1 \}$$

where S is the empirical covariance matrix and ρ the regularization term.

Important points :

- The L_1 regularization impose the sparsity of the precision matrix.
- Null covariance is equivalent to independence under normal distribution.
- Graphical Lasso is a good solution for estimate sparse undirected graphical models.

REFERENCES

Meinshausen, N., Bhlmann, P., 2010. Stability selection. J Roy Stat Soc B 72, 417–473.

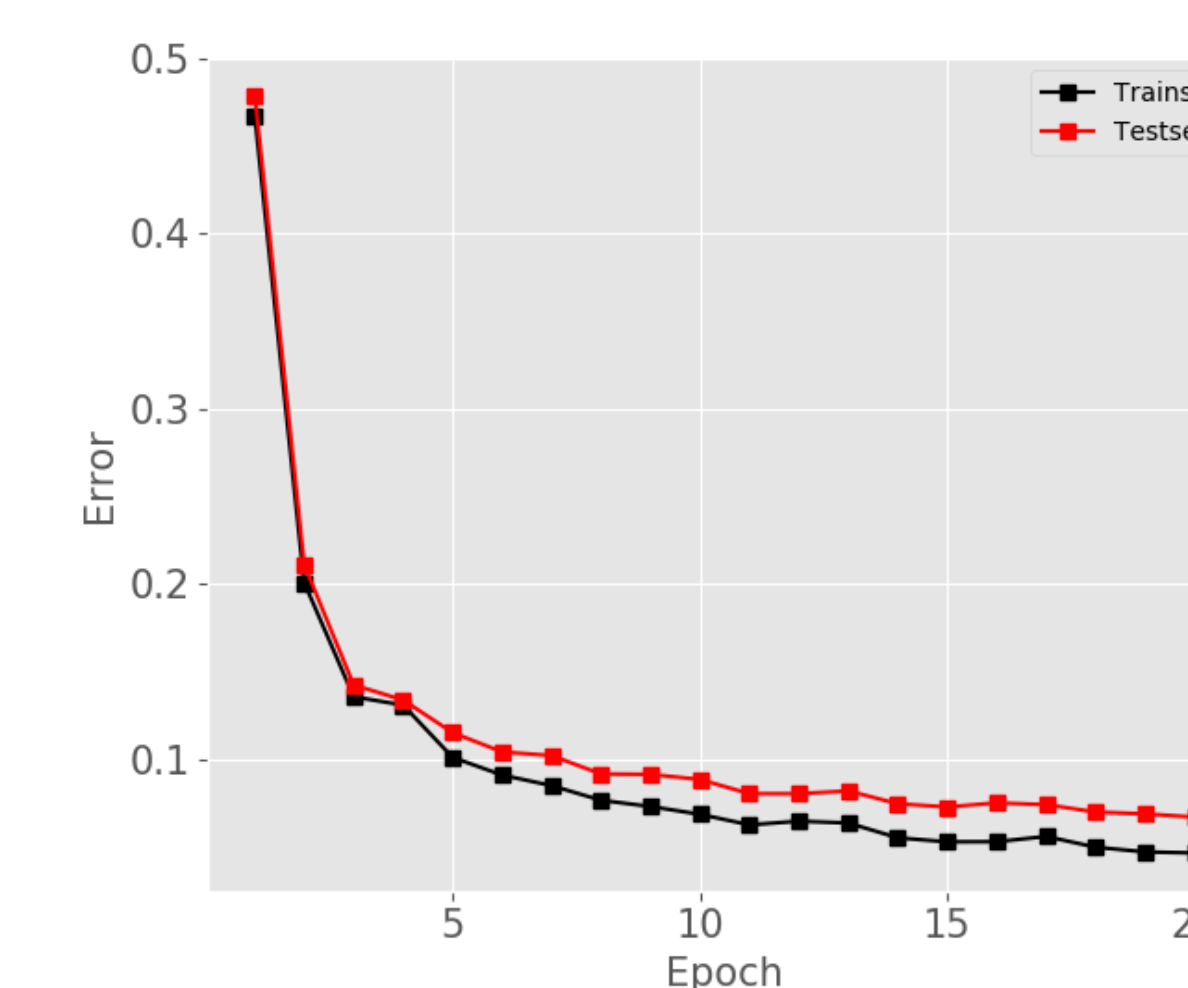
RESULTS

Network Structure

Parameters	Classification	TextGen
Input dim	150k	150k/xx
Embedding dim	128	128
Hidden/cell dim	256	256
Output dim	4	150k/xx
Learning rate	0.0001	0.01
Sequence length	100	100
Batch size	32	16
Dataset size	4×4k	4×4k

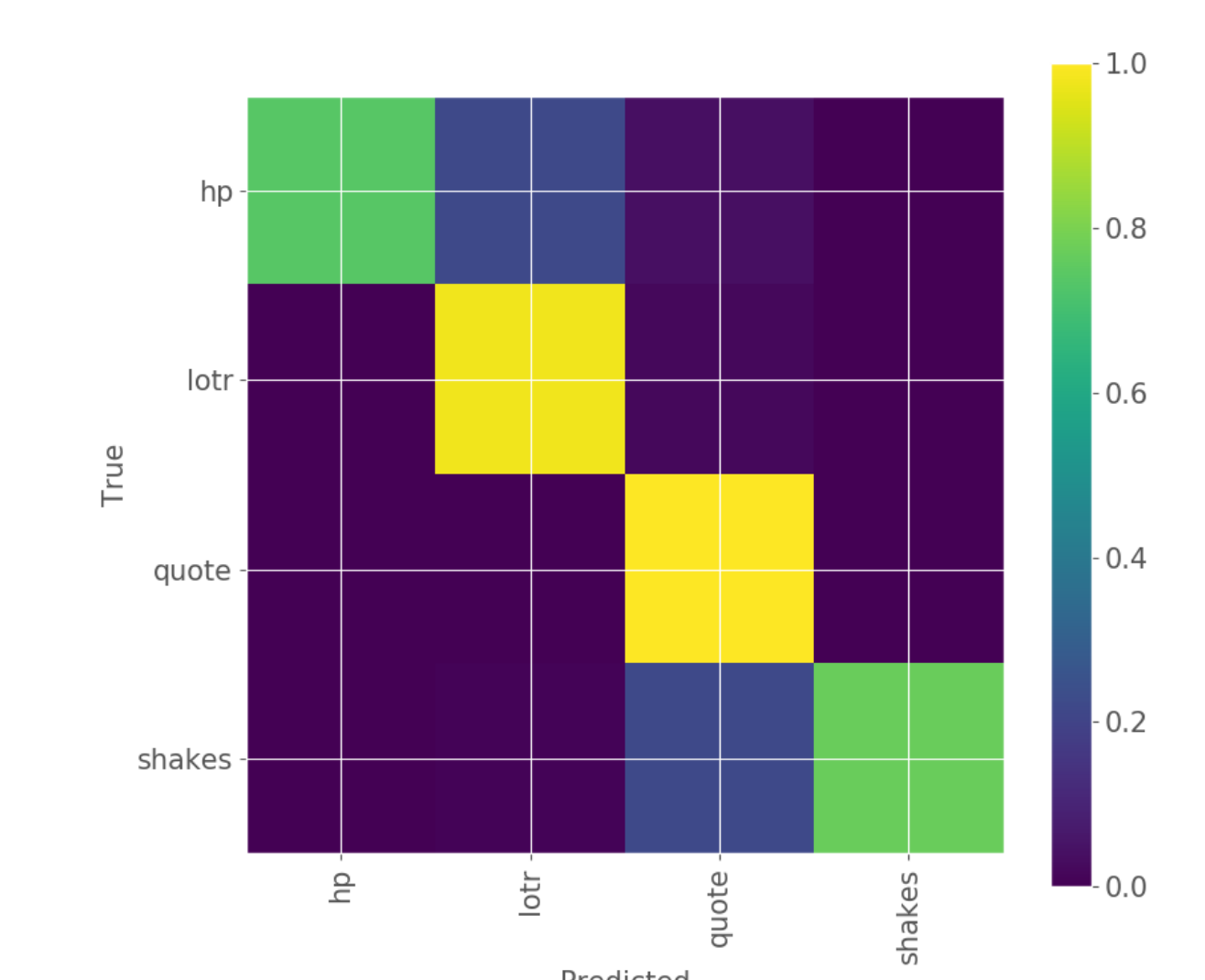
We implemented a LSTM network

Classification of sequences



- 93% fucking beast

TextGen and Classification



- rocking this shit with 87%

Conclusion

- we love you simon
- best project 100%

BLABLABL

- **Comorbidity** : Any distinct additional clinical entity that has existed during the clinical course of a patient.
- **CLSA database** : 50,000 patients and 58 comorbidity factors.
- **Objective** : Profile the burden of comorbidity based on conditional dependences.

Conclusion

- Highlight mechanisms of disease purely based on observational data.
- Focus on few interesting associations which can then be specically tested in model organisms and clinical trials.