University of Stuttgart
Germany

Modeling Sense Structure in Word Usage Graphs with the Weighted Stochastic Block Model

July 21, 2021

Dominik Schlechtweg, Enrique Castaneda,
Jonas Kuhn, Sabine Schulte im Walde
Institute for Natural Language Processing, University of Stuttgart

Introduction

- traditional approach to annotate word senses are binary assignments to sense descriptions (Kilgarriff, 1998)
  - manual effort to create sense descriptions
  - ignores gradedness of word meaning (Erk, McCarthy, & Gaylord, 2013)
- alternative: pairwise semantic proximity judgments of word use pairs (Erk et al., 2013)
  - use pair judgments populate weighted graph (McCarthy, Apidianaki, & Erk, 2016)
  - senses are not annotated directly, but **inferred** on the graph
  - $\rightarrow$ clustering procedure is needed
  - we use the weighted stochastic block model

Data

A	and taking a knife from her pocket, she opened a vein in her little **arm**,
B	And those who remained at home had been heavily taxed to pay for the **arms**, ammunition;
C	and though he saw her within reach of his **arm**, yet the light of her eyes seemed as far off
D	overlooking an **arm** of the sea which, at low tide, was a black and stinking mud-flat
E	twelve miles of coastline lies in the southwest on the Gulf of Aqaba, an **arm** of the Red Sea.
F	when the disembodied **arm** of the Statue of Liberty jets spectacularly out of the

Table 1: Sample of corpus.

Annotation

(A) [. . .] and taking a knife from her pocket, she opened a vein in her little **arm**, and dipping a feather in the blood, wrote something on a piece of white cloth, which was spread before her.

(D) It stood behind a high brick wall, its back windows overlooking an **arm** of the sea which, at low tide, was a black and stinking mud-flat [. . .]

Scale

↑	4: Identical
	3: Closely Related
	2: Distantly Related
	1: Unrelated

Table 2: DURel relatedness scale.

Graph representation

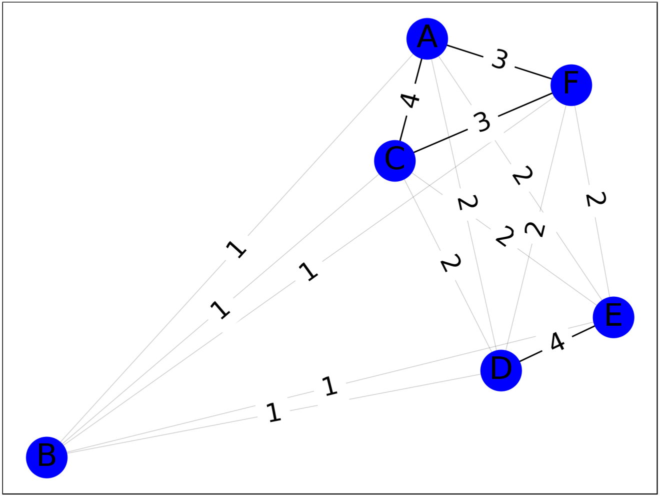

Figure 1: Word Usage Graph of English *arm*. Nodes represent uses of the target word. Edge weights represent the median of proximity judgments between uses.

SemEval WUGs[1]

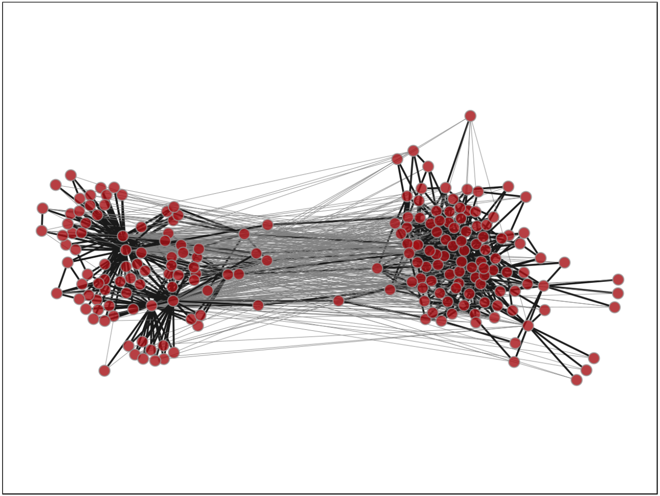

Figure 2: Word Usage Graph of German *zersetzen*.

[1] Schlechtweg, Tahmasebi, Hengchen, Dubossarsky, and McGillivray (2021): `https://www.ims.uni-stuttgart.de/data/wugs`

SemEval WUGs

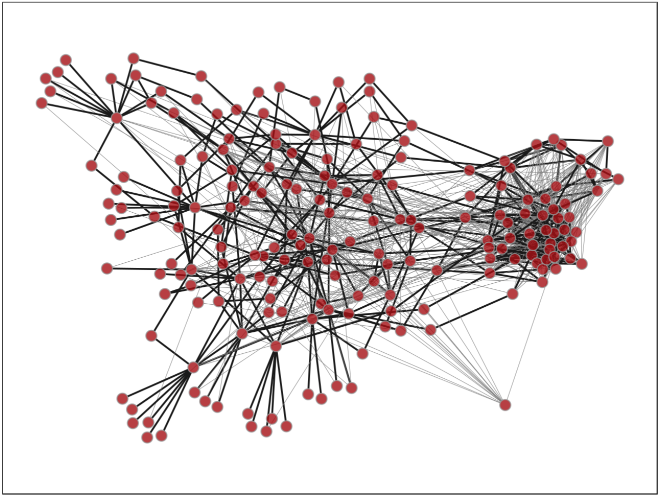

Figure 3: Word Usage Graph of German *Abgesang*.

SemEval WUGs

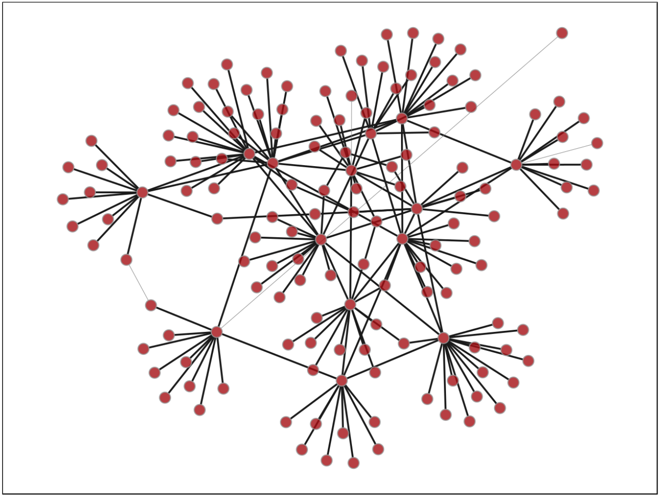

Figure 4: Word Usage Graph of German *Festspiel*.

Weighted Stochastic Block Model (WSBM)

- a generative probabilistic model for random graphs

  (Aicher, Jacobs, & Clauset, 2014; T. P. Peixoto, 2019)
- popular in biology, physics and social sciences
- models nodes as part of blocks (clusters)
- assumes that nodes in the same block are stochastically equivalent
- advantages:
  - allows model selection in absence of ground truth senses
  - captures gradedness by flexible distributions between blocks
  - allows simulation from fitted models
  - extensions allow block (sense) overlap

Inference of Block Structure

- we maximize the Bayesian posterior probability

$$P(b|A,x) = \frac{P(x|A,b)P(A|b)P(b)}{P(A,x)}$$

  where b is the inferred block structure, A is the (unweighted) observed graph, and x are the observed edge weights [2]

  (T. Peixoto, 2017)

- approximation: multilevel agglomerative Markov chain Monte Carlo (T. P. Peixoto, 2014)

[2] All experiments were done with graph-tool: `https://graph-tool.skewed.de/`. Additional code is provided at `https://github.com/kicasta/Modeling_WUGS_WSBM`.

Inferred Structures

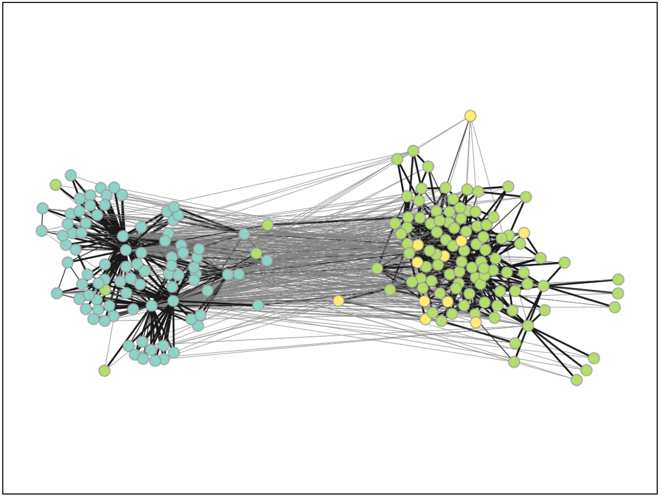

Figure 5: Inferred block structure for *zersetzen*.

Inferred Structures

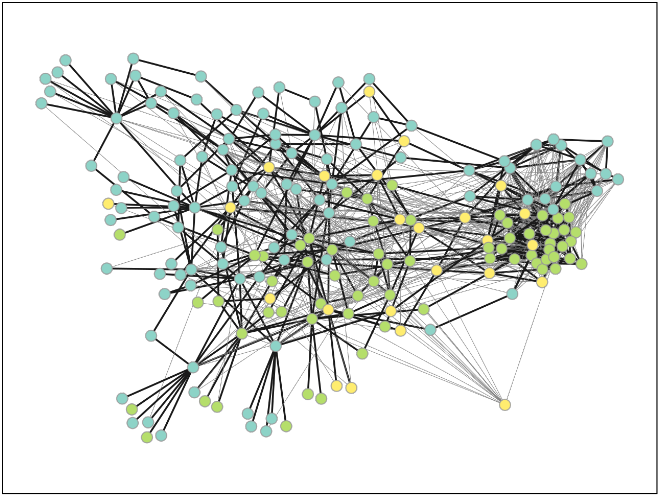

Figure 6: Inferred block structure for *Abgesang*.

Inferred Structures

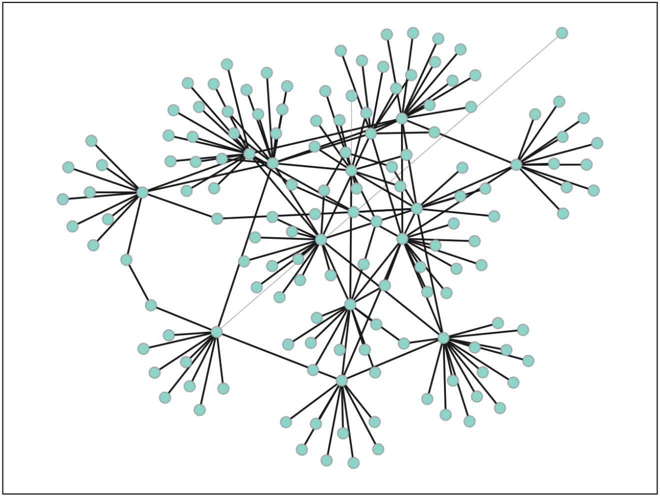

Figure 7: Word Usage Graph for *Festspiel*.

Model Checking – Correspondence to Independent Clustering

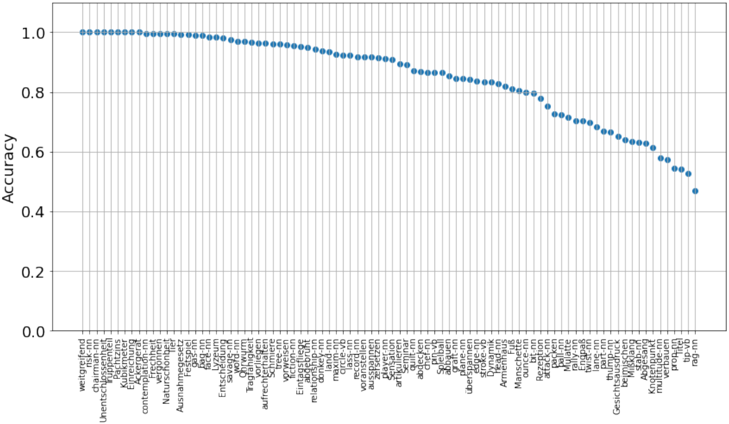

Figure 8: Correspondence to SemEval correlation clustering.

Model Checking – Link Prediction

- how well can a fitted model $P(b|A, x)$ predict weights on masked edges E?
- Inverse Mean Error

$$\text{IME} = 1 - \frac{1}{|E|} \sum_{e \in E} \frac{|e_o - e_p|}{4 - 1}$$

where e_p, e_o correspond to predicted and observed edge weights

Model Checking – Link Prediction

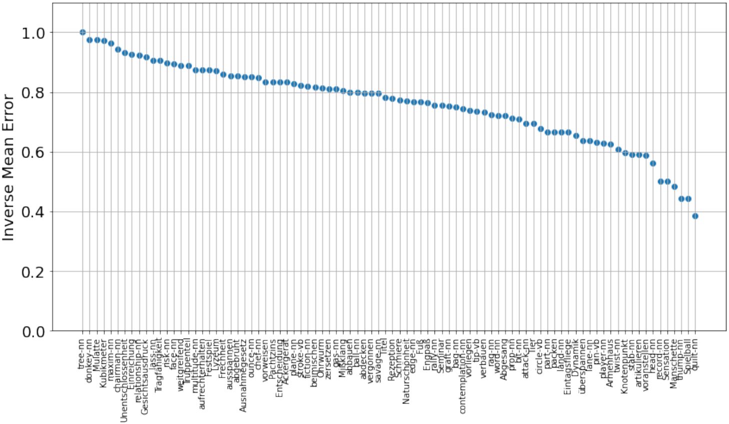

Figure 9: Evaluation result of link prediction.

Model Checking – Predicted/Sampled Graphs

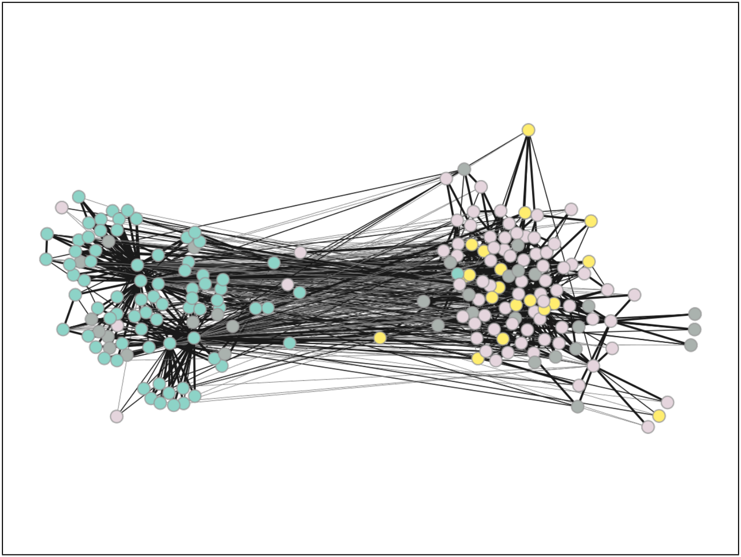

Figure 10: Predicted graph for *zersetzen*.

Model Checking – Predicted/Sampled Graphs

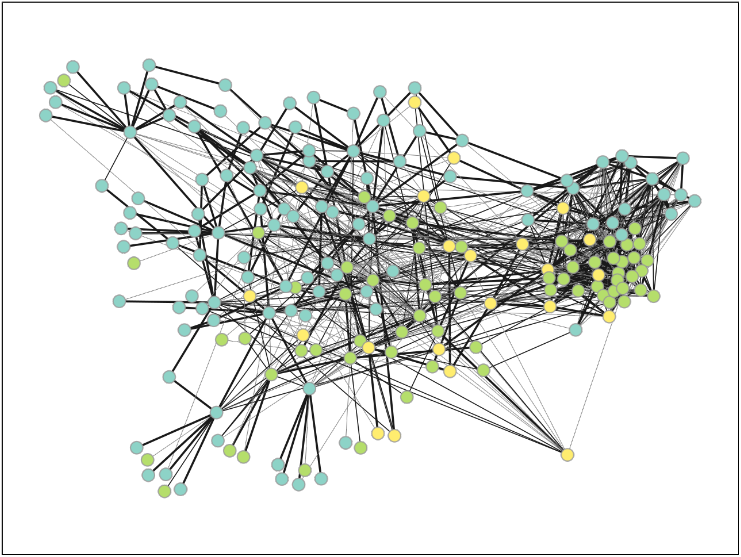

Figure 11: Predicted graph for *Abgesang*.

Model Checking – Predicted/Sampled Graphs

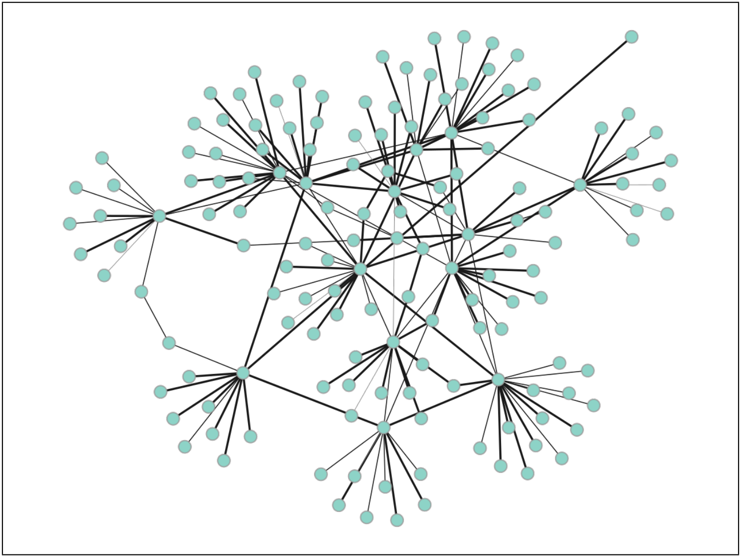

Figure 12: Predicted graph for *Festspiel*.

Model Checking – Fitted Edge Weight Distributions

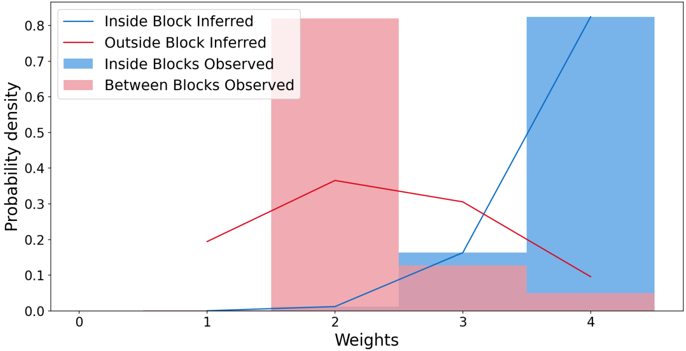

Figure 13: Fitted (line) and observed (bars) edge weight distributions for *zersetzen*.

Model Checking – Fitted Edge Weight Distributions

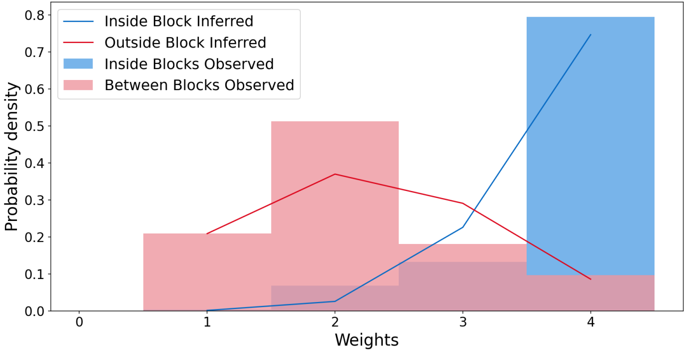

Figure 14: Fitted (line) and observed (bars) edge weight distributions for *Abgesang*.

Model Checking – Fitted Edge Weight Distributions

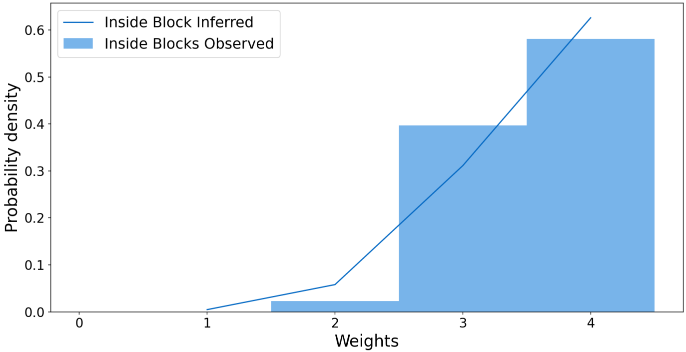

Figure 15: Fitted (line) and observed (bars) edge weight distributions for *Festspiel*.

Conclusion

- we inferred sense structure on WUGs exploiting patterns of semantic proximity
- model selection allows principled inference of sense structures
- the model can be rigorously compared to other probabilistic models (Duda & Hart, 1973; Hoff, Raftery, & Handcock, 2002)
- the inferred structures mostly reflect intuitive sense distinctions
- structural properties of observed graphs are often not very well preserved
  - → more flexible distributions for edge weights are needed
- inferred models can be used for simulation of realistic WUGs[3]
- future: do senses overlap? Which model best describes the data?

[3] `https://www.ims.uni-stuttgart.de/data/wugs`

Bibliography

Aicher, C., Jacobs, A. Z., & Clauset, A. (2014, Jun). Learning latent block structure in weighted networks. *Journal of Complex Networks*, *3*(2), 221—248. Retrieved from `http://dx.doi.org/10.1093/comnet/cnu026` doi: 10.1093/comnet/cnu026

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*.

Erk, K., McCarthy, D., & Gaylord, N. (2013). Measuring word meaning in context. *Computational Linguistics*, *39*(3), 511–554.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, *97*(460), 1090–1098. doi: 10.1198/016214502388618906

Kilgarriff, A. (1998, aug). Senseval: An exercise in evaluating word sense disambiguation programs. In A. M. A. M. S. T. Thierry Fontenelle Philippe Hiligsmann (Ed.), *Proceedings of the 8th euralex international congress* (pp. 167–174). Liège, Belgium: Euralex.

McCarthy, D., Apidianaki, M., & Erk, K. (2016). Word sense clustering and clusterability. *Computational Linguistics*, *42*(2), 245–275.

Peixoto, T. (2017, 08). Nonparametric weighted stochastic block models. *Physical Review E*, *97*. doi: 10.1103/PhysRevE.97.012306

Peixoto, T. P. (2014, Jan). Efficient monte carlo and greedy heuristic for the inference of stochastic block models. *Physical Review E*, *89*(1). Retrieved from `http://dx.doi.org/10.1103/PhysRevE.89.012804` doi: 10.1103/physreve.89.012804

Peixoto, T. P. (2019). Bayesian stochastic blockmodeling. In *Advances in network clustering and blockmodeling* (p. 289-332). John Wiley & Sons, Ltd. doi: 10.1002/9781119483298.ch11

Schlechtweg, D., Tahmasebi, N., Hengchen, S., Dubossarsky, H., & McGillivray, B. (2021). DWUG: A large Resource of Diachronic Word Usage Graphs in Four Languages. *CoRR*, *abs/2104.08540*. Retrieved from `https://arxiv.org/abs/2104.08540`