Supplement for Differential Topic Modeling

Changyou Chen, Wray Buntine, Nan Ding, Lexing Xie, and Lan Du

APPENDIX A
HYBRID GIBBS AND VARIATIONAL METHOD

In this section we propose a hybrid Gibbs and variational inference for our differential topic models. The idea is to first lower bound the model likelihood with variational inference [1], so that it is conjugated to the prior, then use a Gibbs style sampler for this bound.

In this method we will use the usual Chinese restaurant representation of the PYP so that the combinatorial term $\binom{m_{ikw}}{t_{ikw}}^{-1}$ in Equation (4) does not exist. By applying Jensen's inequality first in Equation (4) and then marginalising out $\vec{\phi}^0$, we derive the following lower bound for the likelihood, $p\left(\mathbf{X}, \mathbf{T}, \mathbf{Z} \,\middle|\, \vec{a}, \vec{b}, \vec{\alpha}_{1:I}, \vec{\gamma}, \mathbf{P}\right)$

$$\geq \prod_{k=1}^{K}\left\{\prod_{i=1}^{I}\frac{(b_k|a_k)_{t_{ik\cdot}}}{(b_k)_{m_{ik\cdot}}}\prod_{i=1}^{I}\prod_{w=1}^{V}\prod_{v=1}^{V}\left(\frac{p^i_{wv}}{q^i_{kwv}}\right)^{q^i_{kwv}t_{ikw}}\prod_{i=1}^{I}\prod_{w=1}^{V}S^{m_{ikw}}_{t_{ikw},a_k}\frac{Beta_V\left(\vec{\gamma}+\sum_i\sum_w\vec{q}^{\,i}_{kw}t_{ikw}\right)}{Beta_V\left(\vec{\gamma}\right)}\right\}$$
$$\prod_{i=1}^{I}\prod_{d=1}^{D_i}\frac{Beta_K\left(\vec{\alpha}_i+\vec{n}_{id}\right)}{Beta_K\left(\vec{\alpha}_i\right)}\ , \tag{1}$$

where $\mathbf{T}$ is the table counts t_{ikw} for $i = 1..I, k = 1..K, w = 1..V$, q^i_{kwv}'s are variational variables. This marginalised likelihood bound allows us to develop an approximate algorithm that interleaves a Gibbs sampler over the variables $\mathbf{T}, \mathbf{Z}$ with a variational step to estimate $\vec{q}^{\,i}_{kw}$. The variational step is got by maximising Equation (1) w.r.t. the q^i_{kwv}'s, which results in the following update:

$$q^i_{kwv} \longleftarrow \frac{p^i_{wv}\widehat{\phi^0_{kv}}}{\sum_{v'}p^i_{wv'}\widehat{\phi^0_{kv'}}}\ , \tag{2}$$

C. Chen, W. Buntine and L. Xie are with the Australian National University and National ICT, Australia.

N. Ding is with Google Inc., USA.

L. Du is with the Macquarie University, Australia.

where (using the digamma function $\psi(\cdot)$)

$$\widehat{\phi^0_{kv}} \;=\; e^{\psi\left(\gamma_v+\sum_i \sum_w q^i_{kwv} t_{ikw}\right)} \Bigg/ \sum_v e^{\psi\left(\gamma_v+\sum_i \sum_w q^i_{kwv} t_{ikw}\right)} \quad . \tag{3}$$

A good approximation is simply to use $e^{\psi(x)} \approx x$ in this formula. Based on this, we developed two variational based sampling algorithms for the model, one is to integrate out the base measure $\widehat{\phi^0_{kv}}$, denoted as $\mathtt{Q}$, the other is to retain $\widehat{\phi^0_{kv}}$ while sampling, denoted as $\mathtt{F}$. The resulting psuedo-code are presented in Algorithm 1 and Algorithm 2, respectively.

Algorithm 1 Gibbs-Variational Algorithm $\mathtt{Q}$

1: For each word x^i_{dl} for i, d, l resample the latent topic assignment z^i_{dl} with probabilities proportional to (collecting terms related to $z^i_{dl} = k$) in Equation (1):

$$p(z^i_{dl} = k|\cdot) \;\propto\; \frac{\alpha_{ik} + n_{idk}}{\alpha_{i\cdot} + n_{id\cdot}} \left(\frac{S^{m_{ikw}+1}_{t_{ikw},a_k}}{S^{m_{ikw}}_{t_{ikw},a_k} \cdot (b_k + m_{ik\cdot})} \right)^{\delta_{m_{ikw}>0}}$$
$$\left(\frac{Beta_V\left(\vec{\gamma} + \sum_i \sum_w \vec{q}^{\,i}_{kw} t_{ikw} + q^i_{kwv}\right)}{Beta_V\left(\vec{\gamma} + \sum_i \sum_w \vec{q}^{\,i}_{kw} t_{ikw}\right)} \left(\frac{p^i_{wv}}{q^i_{kwv}}\right)^{q^i_{kwv}} \frac{(b_k + a_k t_{ik\cdot})}{b_k + m_{ik\cdot}} \right)^{\delta_{m_{ikw}\equiv 0}} \quad .$$

2: Resample each table count t_{ikw} for i, k, w by collecting related terms in Equation (1) as:

$$p(t_{ikw}|\cdot) \propto Beta_V\left(\vec{\gamma} + \sum_i \sum_w \vec{q}^{\,i}_{kw} t_{ikw}\right) \left(\frac{p^i_{wv}}{q^i_{kwv}}\right)^{q^i_{kwv}} (b_k|a_k)_{t_{ik\cdot}} S^{m_{ikw}}_{t_{ikw},a_k} \quad .$$

3: For each i, k, w, re-estimate the $\vec{q}^{\,i}_{kw}$ using Equation (2).

APPENDIX B

DEVIATION OF THE FULL LIKELIHOOD FOR THE DIFFERENTIAL TOPIC MODEL

This section shows the detailed derivation of the full likelihood for the proposed differential topic model, using the table indicator representation and auxiliary variables V. First we show that:

Proposition The auxiliary variables v_{ikwt}'s introduced in the main text are valid and reversible, *i.e., marginalizing over these auxiliary variables gets back to the original likelihood.*

Algorithm 2 Gibbs-Variational Algorithm $\mathtt{F}$

1: For each word x^i_{dl} for i, d, l resample the latent topic assignment z^i_{dl} with probabilities proportional to:

$$p(z^i_{dl} = k|\cdot) \;\;\propto\;\; \frac{\alpha_{ik} + n_{idk}}{\alpha_{i\cdot} + n_{id\cdot}} \left(\frac{S^{m_{ikw}+1}_{t_{ikw},a_k}}{S^{m_{ikw}}_{t_{ikw},a_k} \cdot (b_k + m_{ik\cdot})} \right)^{\delta_{m_{ikw}>0}} \left((\sum_v p^i_{kwv} \widehat{\phi^0_{kv}}) \frac{(b_k + a_k t_{ik\cdot})}{b_k + m_{ik\cdot}} \right)^{\delta_{m_{ikw} \equiv 0}} .$$

2: Resample each table count t_{ikw} for i, k, w as:

$$p(t_{ikw}|\cdot) \propto \left(\sum_v p^i_{kwv} \widehat{\phi^0_{kv}} \right)^{t_{ikw}} (b_k|a_k)_{t_{ik\cdot}} S^{m_{ikw}}_{t_{ikw},a_k} .$$

3: For each i, k, w, re-estimate the $\vec{q}^{\,i}_{kw}$ using Equation (2).

Proof:

To see this, using the same notation, from Equation (4) in the main text we have

$$\begin{aligned}
&P_r(\vec{x}_{1:I,1:D_i}, \vec{t}_{1:I,1:K_i}, \vec{z}_{1:I,1:D_i} | \vec{a}, \vec{b}, \vec{\alpha}_{1:I}, \vec{\gamma}, \mathbf{P}^{1:I}) \\
&= \prod_k \left\{ \prod_{i=1}^I \frac{(b_k|a_k)_{t_{ik\cdot}}}{(b_k)_{m_{ik\cdot}}} \prod_{i=1}^I \prod_w \left(\sum_v p^i_{wv} \phi^0_{kv} \right)^{t_{ikw}} \right. \\
&\quad \left. \prod_{i=1}^I \prod_w S^{m_{ikw}}_{t_{ikw},a_k} \frac{\Gamma(\sum_v \gamma_v)}{\prod_v \Gamma(\gamma_v)} \prod_v \left(\phi^0_{kv}\right)^{\gamma_v - 1} \right\} \prod_{i=1}^I \prod_{d=1}^{D_i} \frac{Beta_K\left(\vec{\alpha_i} + \vec{n}_{id}\right)}{Beta_K\left(\vec{\alpha_i}\right)} \\
&= \prod_k \left\{ \prod_{i=1}^I \frac{(b_k|a_k)_{t_{ik\cdot}}}{(b_k)_{m_{ik\cdot}}} \prod_{i=1}^I \prod_w \prod_{t=1}^{t_{ikw}} \left(\sum_v p^i_{wv} \phi^0_{kv} \right) \right. \\
&\quad \left. \prod_{i=1}^I \prod_w S^{m_{ikw}}_{t_{ikw},a_k} \frac{\Gamma(\sum_v \gamma_v)}{\prod_v \Gamma(\gamma_v)} \prod_v \left(\phi^0_{kv}\right)^{\gamma_v - 1} \right\} \prod_{i=1}^I \prod_{d=1}^{D_i} \frac{Beta_K\left(\vec{\alpha_i} + \vec{n}_{id}\right)}{Beta_K\left(\vec{\alpha_i}\right)} \\
&= \prod_k \left\{ \prod_{i=1}^I \frac{(b_k|a_k)_{t_{ik\cdot}}}{(b_k)_{m_{ik\cdot}}} \prod_{i=1}^I \prod_w \prod_{t=1}^{t_{ikw}} \left(\sum_{v_{ikwt}} p^i_{wv_{ikwt}} \phi^0_{kv_{ikwt}} \right) \right. \\
&\quad \left. \prod_{i=1}^I \prod_w S^{m_{ikw}}_{t_{ikw},a_k} \frac{\Gamma(\sum_v \gamma_v)}{\prod_v \Gamma(\gamma_v)} \prod_v \left(\phi^0_{kv}\right)^{\gamma_v - 1} \right\} \prod_{i=1}^I \prod_{d=1}^{D_i} \frac{Beta_K\left(\vec{\alpha_i} + \vec{n}_{id}\right)}{Beta_K\left(\vec{\alpha_i}\right)}
\end{aligned}$$

$$= \prod_k \left\{ \prod_{i=1}^{I} \prod_w \left\{ \sum_{v_{ikw1}=1}^{V} \cdots \sum_{v_{ikwt_{ikw}}=1}^{V} \left(p^i_{wv_{ikw1}} \phi^0_{kv_{ikw1}}\right) \cdots \left(p^i_{wv_{ikwt_{ikw}}} \phi^0_{kv_{ikwt_{ikw}}}\right) \right\} \right. \tag{4}$$
$$\left. \prod_{i=1}^{I} \frac{(b_k|a_k)_{t_{ik\cdot}}}{(b_k)_{m_{ik\cdot}}} \prod_{i=1}^{I} \prod_w S^{m_{ikw}}_{t_{ikw},a_k} \frac{\Gamma(\sum_v \gamma_v)}{\prod_v \Gamma(\gamma_v)} \prod_v \left(\phi^0_{kv}\right)^{\gamma_v - 1} \right\} \prod_{i=1}^{I} \prod_{d=1}^{D_i} \frac{Beta_K\left(\vec{\alpha}_i + \vec{n}_{id}\right)}{Beta_K\left(\vec{\alpha}_i\right)}$$

Now consider the summation term in the above formula with index k dropped

$$\prod_{i=1}^{I} \prod_w \left\{ \sum_{v_{iw1}=1}^{V} \cdots \sum_{v_{iwt_{iw}}=1}^{V} \left(p^i_{wv_{iw1}} \phi^0_{v_{iw1}}\right) \cdots \left(p^i_{wv_{iwt_{iw}}} \phi^0_{v_{iwt_{iw}}}\right) \right\}, \tag{5}$$

we use i' to index (i, w), j to index $(v_{iw1}, \cdots, v_{iw(t_{iw}-1)})$, and k' to index $(v_{iwt_{iw}})$, *i.e.*,

$$\prod_{i'} \triangleq \prod_i \prod_w$$
$$\sum_j \triangleq \sum_{v_{iw1}} \cdots \sum_{v_{iw(t_{iw}-1)}}$$
$$\sum_{k'} \triangleq \sum_{v_{iwt_{iw}}}$$

where $i' = 1, \cdots, I$, $j = 1, \cdots, J$ and $k' = 1, \cdots, K$. Furthermore, we use the following simplified notation:

$$f_{i'j} \triangleq \left(p^i_{wv_{iw1}} \phi^0_{v_{iw1}}\right) \cdots \left(p^i_{wv_{iw(t_{iw}-1)}} \phi^0_{v_{iw(t_{iw}-1)}}\right)$$
$$g_{i'jk'} \triangleq p^i_{wv_{iwt_{iw}}} \phi^0_{v_{iwt_{iw}}}$$

Then (5) can be simplified as:

$$\begin{aligned}
(5) &= \prod_{i'} \sum_j \sum_{k'} f_{i'j} g_{i'jk'} \\
&= \prod_{i'} \sum_{k'} \left(\sum_j f_{i'j} g_{i'jk'} \right) \\
&= \sum_{k_1} \left(\sum_j f_{1j} g_{1jk_1} \right) \sum_{k_2} \left(\sum_j f_{2j} g_{2jk_2} \right) \cdots \sum_{k_I} \left(\sum_j f_{Ij} g_{Ijk_I} \right) \\
&= \sum_{k_1} \cdots \sum_{k_I} \left(\prod_{i'} \left(\sum_j f_{i'j} g_{i'jk_i} \right) \right) \\
&= \sum_{k'} \prod_{i'} \left(\sum_j f_{i'j} g_{i'jk_i} \right)
\end{aligned} \tag{6}$$

This means we can swap the summation out of the product without changing the formula of (5), by induction we conclude that the summation terms in (4) can be swapped out of the multiplication terms. Using $p(\cdots)$ to denote the original likelihood without auxiliary variables, (6) means that

$$p(\cdots) = \sum_{v_{ikw1}} \cdots \sum_{v_{ikwt_{ikw}}} p(\cdots, v_{ikw1}, \cdots, v_{ikwt_{ikw}}).$$

Thus the variables $(v_{ikw1}, \cdots, v_{ikwt_{ikw}})$ in (4) are valid auxiliary variables, they are reversible. By using these auxiliary variables, the power of a summation term can be simplified as product terms.

■

Now we are ready to use these auxiliary variables to derive the augmented likelihood. Let $\vec{V} = \{v_{ikwt}, \forall i, k, w, t\}$ and $q^i_{kwv} = \sum_{t=1}^{t_{ikw}} 1_{v_{ikwt}=v}$, $Q = \{q^i_{kwv}, \forall i, k, w, v\}$, and $R = \{r_{idl}\}$ be the table indicator variables, then

$$\begin{aligned}
& P_r(\vec{x}_{1:I,1:D_i}, \vec{z}_{1:I,1:D_i}, \vec{V}, \vec{R} | \vec{a}, \vec{b}, \vec{\alpha}_{1:I}, \vec{\gamma}, \mathbf{P}^{1:I}) \\
\propto\ & \left(\prod_i \prod_k \prod_w \frac{t_{ikw}!(m_{ikw} - t_{ikw})!}{m_{ikw}!} \right) \prod_k \left\{ \prod_{i=1}^I \frac{(b_k|a_k)_{t_{ik\cdot}}}{(b_k)_{m_{ik\cdot}}} \prod_{i=1}^I \prod_w \prod_{t=1}^{t_{ikw}} \left(p^i_{wv_{ikwt}} \phi^0_{kv_{ikwt}} \right) \right. \\
& \left. \prod_{i=1}^I \prod_w S^{m_{ikw}}_{t_{ikw}, a_k} \frac{\Gamma(\sum_v \gamma_v)}{\prod_v \Gamma(\gamma_v)} \prod_v \left(\phi^0_{kv} \right)^{\gamma_v - 1} \right\} \prod_{i=1}^I \prod_{d=1}^{D_i} \frac{Beta_K\left(\vec{\alpha}_i + \vec{n}_{id}\right)}{Beta_K\left(\vec{\alpha}_i\right)} \\
=\ & \left(\prod_i \prod_k \prod_w \frac{t_{ikw}!(m_{ikw} - t_{ikw})!}{m_{ikw}!} \right) \prod_{i=1}^I \prod_w \prod_v \left(p^i_{wv} \right)^{\sum_k q^i_{kwv}} \prod_k \left\{ \prod_{i=1}^I \frac{(b_k|a_k)_{t_{ik\cdot}}}{(b_k)_{m_{ik\cdot}}} \right. \\
& \left. \prod_v \left(\phi^0_{kv} \right)^{\sum_i \sum_w q^i_{kwv} + \gamma_v - 1} \prod_{i=1}^I \prod_w S^{m_{ikw}}_{t_{ikw}, a_k} \frac{\Gamma(\sum_v \gamma_v)}{\prod_v \Gamma(\gamma_v)} \right\} \prod_{i=1}^I \prod_{d=1}^{D_i} \frac{Beta_K\left(\vec{\alpha}_i + \vec{n}_{id}\right)}{Beta_K\left(\vec{\alpha}_i\right)} \\
& \xrightarrow{\text{Integrate out } \phi^0} \\
& \left(\prod_i \prod_k \prod_w \frac{t_{ikw}!(m_{ikw} - t_{ikw})!}{m_{ikw}!} \right) \prod_{i=1}^I \prod_w \prod_v \left(p^i_{wv} \right)^{\sum_k q^i_{kwv}} \prod_{i=1}^I \prod_{d=1}^{D_i} \frac{Beta_K\left(\vec{\alpha}_i + \vec{n}_{id}\right)}{Beta_K\left(\vec{\alpha}_i\right)} \\
& \prod_k \left\{ \prod_{i=1}^I \frac{(b_k|a_k)_{t_{ik\cdot}}}{(b_k)_{m_{ik\cdot}}} \frac{\prod_v \Gamma\left(\sum_i \sum_w q^i_{kwv} + \gamma_v\right)}{\Gamma\left(\sum_i \sum_w t_{ikw} + \sum_v \gamma_v\right)} \prod_{i=1}^I \prod_w S^{m_{ikw}}_{t_{ikw}, a_k} \frac{\Gamma\left(\sum_v \gamma_v\right)}{\prod_v \Gamma\left(\gamma_v\right)} \right\} . \qquad (7)
\end{aligned}$$

APPENDIX C
EXTRA EXPERIMENT RESULTS

A. Illustration

Fig. 1 illustrates Table 1 in the main text to show the probabilistic graphical model on word vectors for the topic hierarchy learned on a Reuters News dataset `GENT` consisting 6 groups. Note that we did not threshold the b values thus all groups have the same number of topics paired together. We can easily see how topics change over different groups in the figure.

B. Topic Alignment

To investigate the topic alignment ability of the proposed topic model, we give some quantitative results here which compares the topic pairing in `TII` with the same topics but aligned by using the Hellinger distance [2]. `TI` is not used here because it modifies the (simple Hellinger) distances due to lexical cohesion. We wrote a pairwise "topic alignment" module and applied it to the `TII` results for `MLJ`, `GDIS` and `GENT`. These gave 100%, 72% and 97% agreement respectively on average with the paired topics produced by `TII`. Note this is less for `GDIS` because there are more non-paired topics created for this dataset. Thus `TII` produces excellent topic alignments that is as good as those that might be produced *post hoc*. This was also reflected in Fig. 4 in the main text. Moreover, the Hellinger distance between aligned topics in the TII model and the model standard deviation for the topics of $\sqrt{\frac{1-a_k}{1+b_k}}$ are strongly related. They have a Pearson correlation coefficient of 0.792 with a slope close to 1.0. Thus the model is effective at producing well aligned sets of topics.

C. Word Association

Figure 2 below gives a more completed illustration of the *word association* other than the one given in the main text.

D. Detailed Comparison

We first restate the algorithms and notation we use in the experiments. We have 8 algorithms to compare, which are

- `TI`: the full Gibbs table indicator sampler for the TPYP.

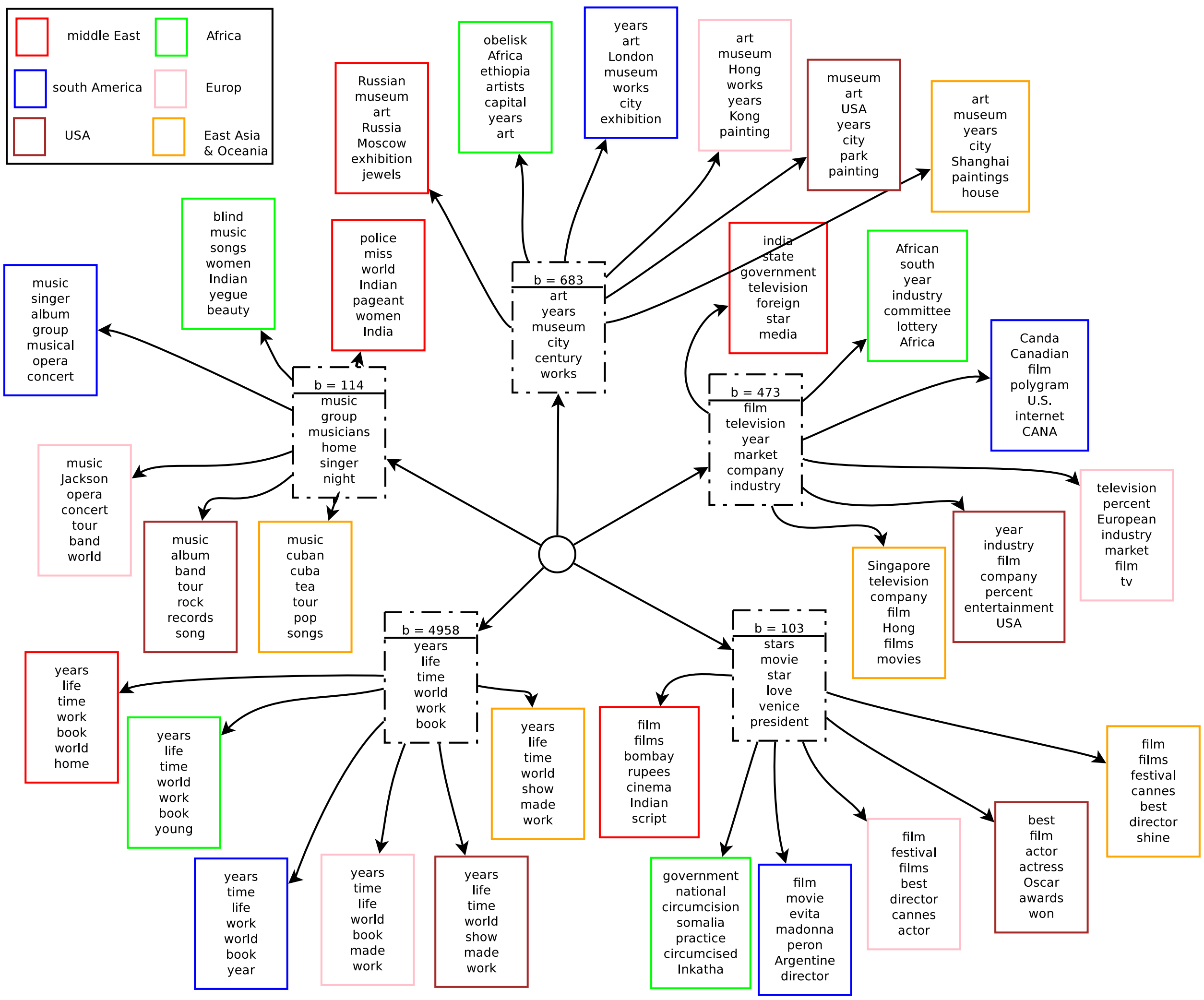


Fig. 1: An example of a topic hierarchy, a probabilistic graphical model on word vectors, learned for an Entertainment News dataset extracted from Reuters 1996-1997 articles, described later denoted as GENT. For the dataset ran with 10 topics (5 plotted), the boxes with dash line are the master topics, while the colored boxes correspond to topics in the six groups. Values of b given represent concentration (inverse variance), reflecting the variation amongst children topics across regions. Best viewed in color.

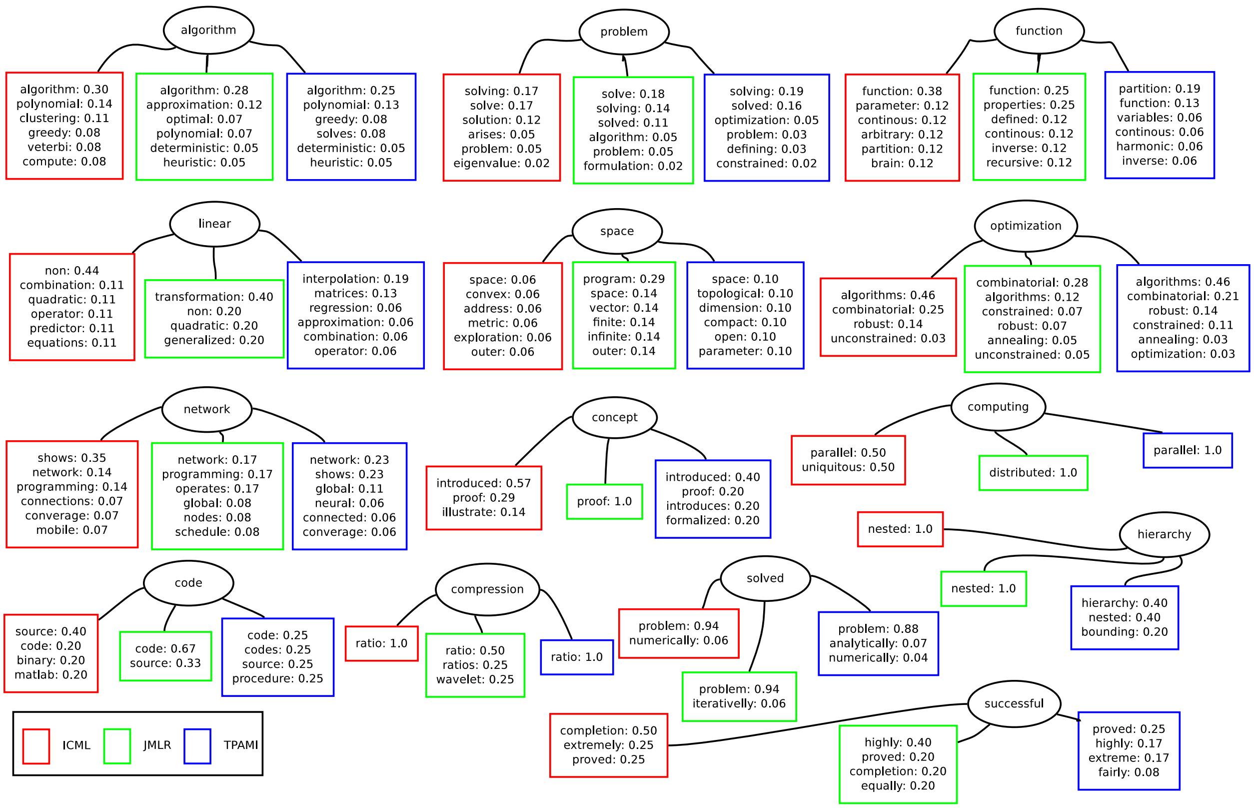


Fig. 2: An example of word association structure learned by the model on `MLJ` dataset. The words in the eclipses are from the global vocabulary, each of them corresponds to a set of words (in the colored boxes) in each group, represented by the statistics $\{\tilde{q}^i_{kwv}\}$ derived from *word association* in eq.(7) in the main text, the numbers following the words represent the strength of the correlations in range $[0, 1]$. Best viewed in color.

- `Q`: the hybrid Gibbs and variational method proposed in Appendix A.
- `F`: a variant of `Q` by keeping the approximated ϕ^0 instead of integrating it out as in (1) in Appendix A. We used this variant because we found it performs better than `Q`.
- `TII`: a degenerated `TI` with identity transformation matrix I.
- `CS`: the collapsed Gibbs sampler for the hierarchical `PDP` [3].
- `SS`: a variant of the CRP based algorithm, originally the `sampling by direct assignment` algorithm proposed for the HDP [4].
- `PDP`: use `PDP` as the prior for the topic-word distributions for each group separately, equivalent to [5].

- `LDA`: plain LDA [6] trained on each group.

Note that the first three algorithms deal with non-identity transformation matrix, thus have considered word correlation information into the model, while the last five do not. Since we construct the transformation matrix in two ways, we will use subscripts 'co' and 'wn' to denote the algorithms using the matrices constructed from Wikipedia and WordNet, respectively. All the algorithms are run using 2000 Gibbs/variational cycles as burn in, which is adequate for convergence in the experiments, and 100 samples are collected for perplexity calculation. The hyperpameters are also sampled during inference, but with the *discount parameter* a set to 0.7, known to perform best in topic-word distribution modeling.

Figure 3 summaries the comparison results. The finding is consistent with the main text. Interestingly, we find that the hybrid Gibbs and variational methods `Q` and `F` fail to compete with other algorithms, sometimes even not as well as LDA.

E. Handwritten Digits Modeling

Other than the three group illustration given in Fig.9 in the main text, here we given a complete figure which illustrates different and sharing structures between all the groups in Figure 4.

Acknowledgment

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Center of Excellence program.

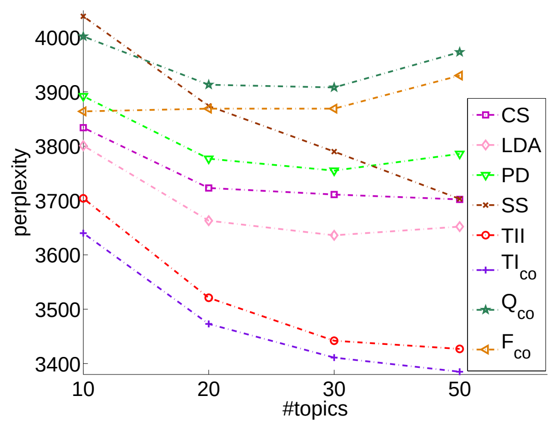
References

[1] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Mach. Learn.*, vol. 37, pp. 183–233, 1999.

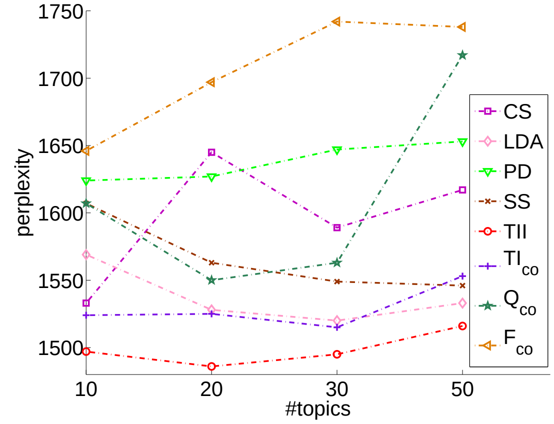
[2] J. Petterson, A. Smola, T. Caetano, W. Buntine, and S. Narayanamurthy, "Word features for latent Dirichlet allocation," in *NIPS*, 2010.

[3] L. Du, W. Buntine, and H. Jin, "A segmented topic model based on the two-parameter Poisson-Dirichlet process," *Machine Learning*, vol. 81, pp. 5–19, 2010.

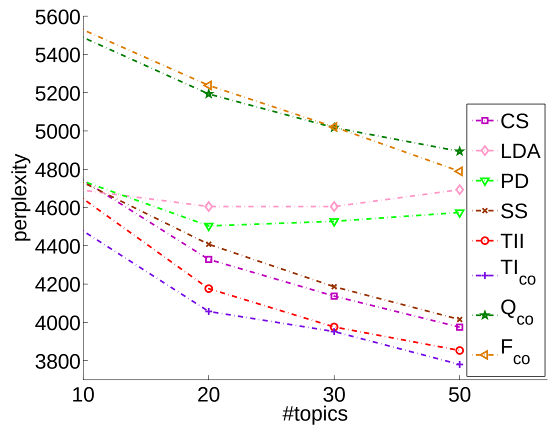
[4] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *Journal of the ASA*, vol. 101, no. 476, pp. 1566–1581, 2006.



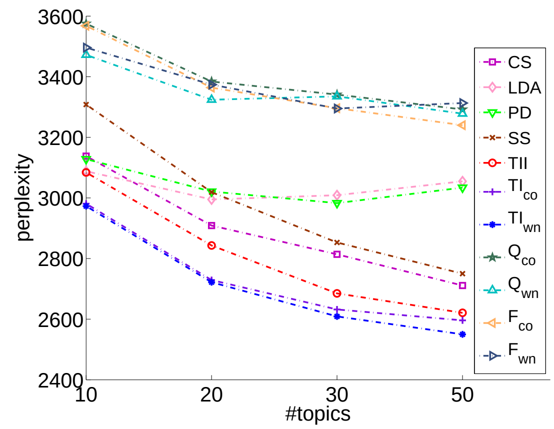
(a) BD



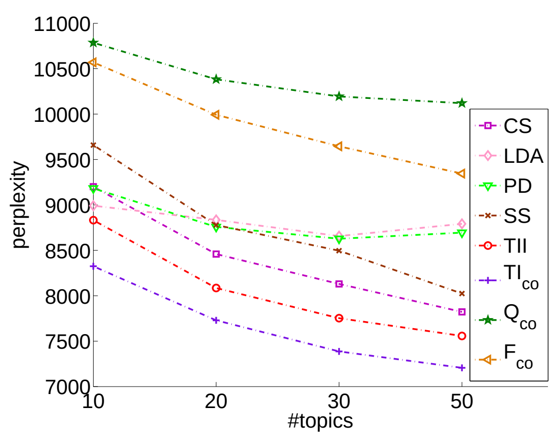
(b) MLJ



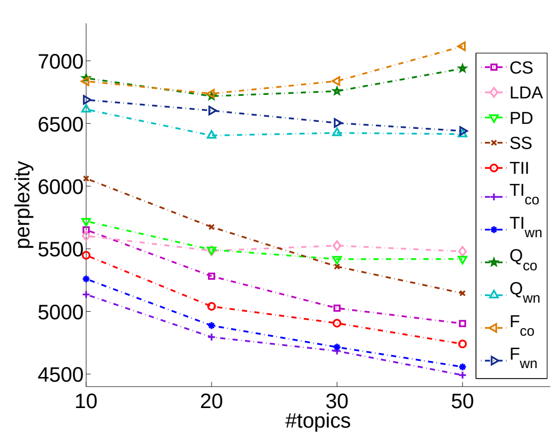
(c) gdis_CC



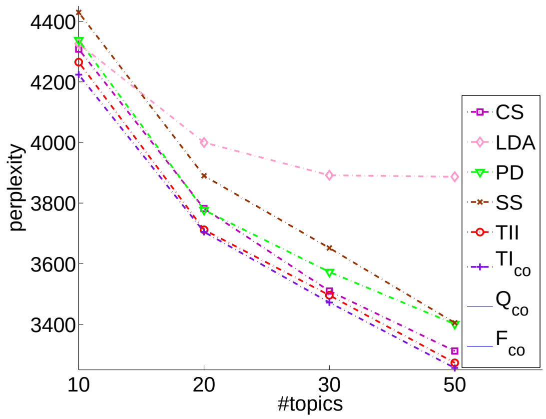
(d) gdis_CCp2



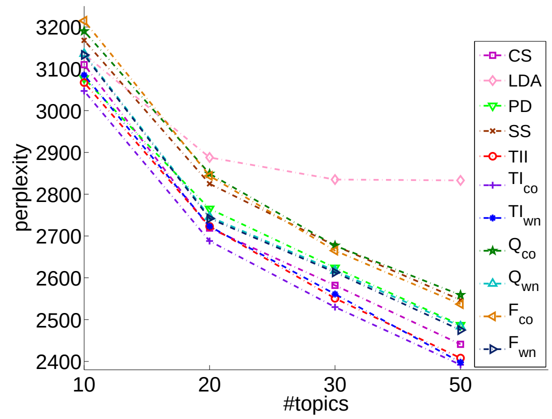
(e) gent_CC



(f) gent_CCp2



(g) gpol_CC



(h) gpol_CCp2

Fig. 3: Results on the eight datasets with all algorithms and different transformation matrices (indicated by `wn` and `co`). Our model performs best. Best viewed in color.

[5] I. Sato and H. Nakagawa, "Topic models with power-law using Pitman-Yor process," in *SIGKDD*, 2010.

[6] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

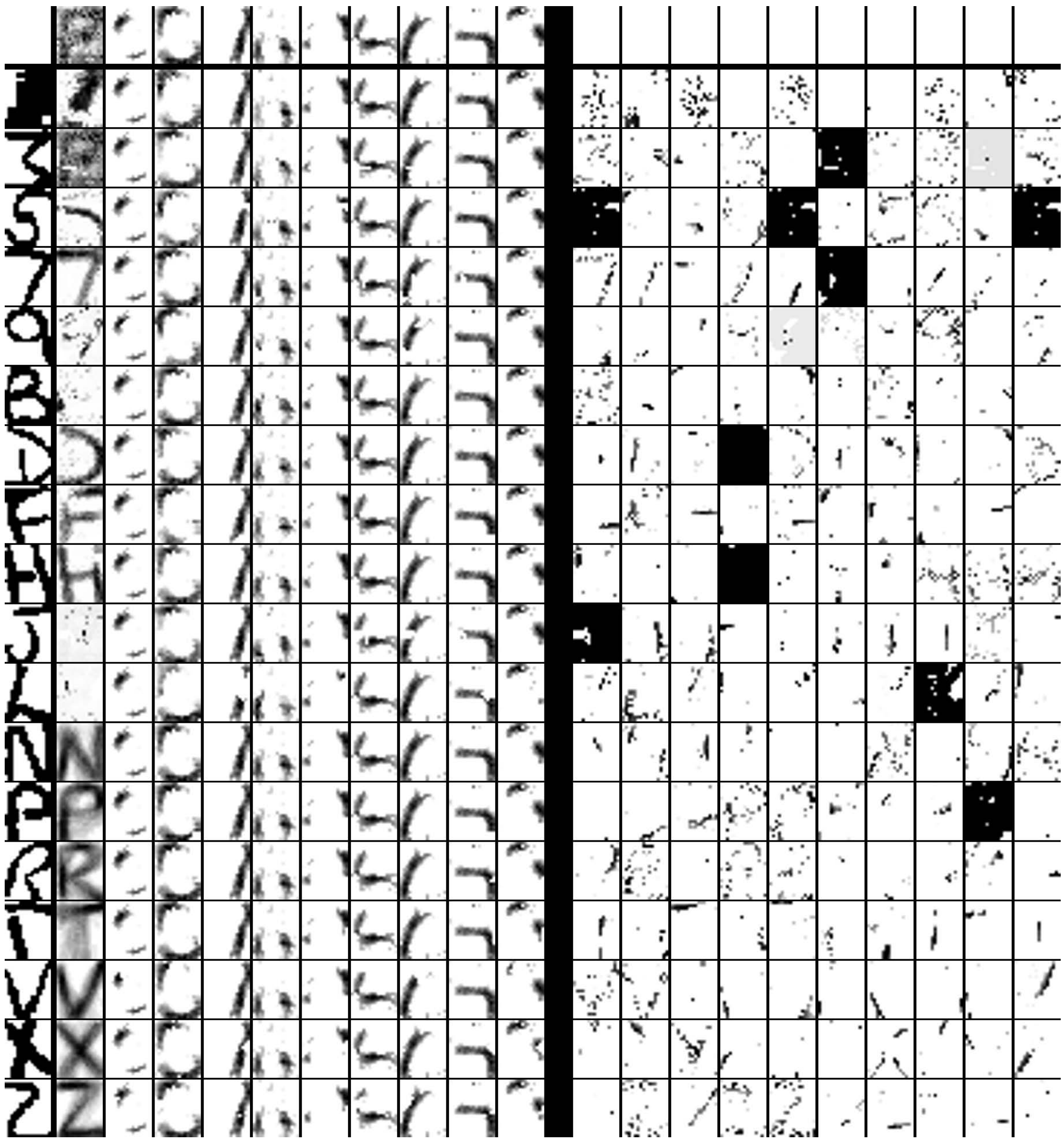


Fig. 4: Results on `BinaryAlphaDigs` datasets for TII (left) and LDA (right) with 20 topics (10 shown). The first column contains a random sample from the dataset for each group. The other columns represent the 10 topics for each group. The first row of TII topics represents the master (parent) topics. It is interesting to see that the second column of TII topics reveals different structures among the digits/characters while the other columns represent the similar structures.