Supplementary Material for "Inter and Intra Topic Structure Learning with Word Embeddings"

He Zhao [1] Lan Du [1] Wray Buntine [1] Mingyaun Zhou [2]

1. Inference Details

Recall that the data is the word count vector $\boldsymbol{x}_j^{(1)}$ of document j. Given the PFA framework, we can apply the collapsed Gibbs sampling to sample the bottom-layer topic for the i-th word in j, v_{ji}, similar to Eq. (28) in Appendix B of Zhou et al. (2016), as follows:

$$P(z_{ji} = k_1) \propto \frac{\beta_{vk_1} + x_{v_{ji}\cdot k_1}^{(1)^{-ji}}}{\beta_{\cdot k_1} + x_{\cdot\cdot k_1}^{(1)^{-ji}}} \left(x_{\cdot jk_1}^{(1)^{-ji}} + \boldsymbol{\phi}_{k_1:}^{(2)} \boldsymbol{\theta}_j^{(2)} \right) \tag{1}$$

where z_{ji} is the topic index for v_{ji} and $x_{vjk}^{(1)} := \sum_i \delta(v_{ji} = v, z_{ji} = k_1)$ counts the number of times that term v appears in document j; we use x^{-ji} to denote the count x calculated without considering word i in document j.

Given the latent counts $x_{v\cdot k_1}^{(1)}$, the multinomial likelihood of $\boldsymbol{\phi}_{k_1}^{(1)}$ is proportional to $\left(\phi_{vk_1}^{(1)}\right)^{x_{v\cdot k_1}^{(1)}}$. Due to the Dirichlet-multinomial conjugacy, the joint likelihood of $\boldsymbol{\beta}_{k_1}$ is computed as:

$$\mathcal{L}\left(\beta_{vk_1}\right) \propto \frac{\Gamma(\beta_{\cdot k_1})}{\Gamma(\beta_{k_1 \cdot} + x_{\cdot\cdot k_1}^{(1)})} \prod_v^V \frac{\Gamma(\beta_{vk_1} + x_{v\cdot k_1}^{(1)})}{\Gamma(\beta_{vk_1})}. \tag{2}$$

By introducing an auxiliary beta distributed variable:

$$q_{k_1} \sim \text{Beta}(\beta_{\cdot k_1}, x_{\cdot\cdot k_1}^{(1)}), \tag{3}$$

we can transform the first gamma ratio in Eq. (2) to $(q_{k_1})^{\beta_{\cdot k_1}}$ (Zhao et al., 2018). After this augmentation, one can show that the likelihood of $\beta_{k_1,v}$ is proportional to the negative binomial distribution.

Recall that $\beta_{vk_1} = \sum_s^S \beta_{vk_1}^{<s>}$ and the shape parameter of $\beta_{vk_1}^{<s>}$, $\alpha_{k_1}^{<s>}$ is drawn from hierarchical gamma. This construction is closely related to the gamma-negative binomial process (Zhou & Carin, 2015; Zhou, 2016), which enables the model to automatically determine the number of effective sub-topics.

Furthermore, we introduce another auxiliary variable:

$$h_{vk_1} \sim \text{CRT}\left(x_{v\cdot k_1}^{(1)}, \beta_{vk_1}\right), \tag{4}$$

where $h \sim \text{CRT}(n, r)$ stands for the Chinese Restaurant Table distribution (Zhou & Carin, 2015) that generates the number of tables h seated by n customers in a Chinese restaurant process with the concentration parameter r (Wray & Marcus, 2012). Given h_{vk_1}, the second gamma ratio can be augmented as $(\beta_{vk_1})^{h_{vk_1}}$. Finally, with the two auxiliary variables, Eq. (2) can be written as:

$$\mathcal{L}\left(\beta_{vk_1}, q_{k_1}, h_{vk_1}\right) \propto (q_{k_1})^{\beta_{vk_1}} (\beta_{vk_1})^{h_{vk_1}}. \tag{5}$$

Sample $\beta_{vk_1}^{<s>}$. Given the table counts h_{vk_1} in Eq. (5), we can sample the counts $h_{vk_1}^{<s>}$ for each sub-topic s as follows:

$$\left(h_{vk_1}^{<1>}, \cdots, h_{vk_1}^{<S>}\right) \sim \text{Mult}\left(h_{vk_1}, \frac{\beta_{vk_1}^{<1>}}{\beta_{vk_1}}, \cdots, \frac{\beta_{vk_1}^{<S>}}{\beta_{vk_1}}\right). \tag{6}$$

Given $h_{vk_1}^{<s>}$, we can sample $\beta_{vk_1}^{<s>}$ as:

$$\beta_{vk_1}^{<s>} \sim \frac{\text{Gam}(\alpha_{k_1}^{<s>} + h_{vk_1}^{<s>}, 1)}{e^{-\pi_{vk_1}^{<s>}} + \log \frac{1}{q_{k_1}}}, \tag{7}$$

where we define

$$\pi_{vk_1}^{<s>} := \boldsymbol{f}_v^\top \boldsymbol{w}_{k_1}^{<s>}. \tag{8}$$

Sample $\alpha_{k_1}^{<s>}$. By integrating $\beta_{vk_1}^{<s>}$ out, we sample $\alpha_{k_1}^{<s>}$ as:

$$\alpha_{k_1}^{<s>} \sim \frac{\text{Gam}(\alpha_0^{<s>}/S + g_{\cdot k_1}^{<s>}, 1)}{c_0^{<s>} + \log\left(1 + e^{\pi_{vk_1}^{<s>}} \log \frac{1}{q_{k_1}}\right)}, \tag{9}$$

where

$$g_{\cdot k_1}^{<s>} := \sum_v^V g_{vk_1}^{<s>}, \tag{10}$$

$$g_{vk_1}^{<s>} \sim \text{CRT}\left(h_{vk_1}^{<s>}, \alpha_{k_1}^{<s>}\right). \tag{11}$$

According to the gamma-gamma conjugacy, $\alpha_0^{<s>}$ and $c_0^{<s>}$ can be sampled in a similar way.

[1] Faculty of Information Technology, Monash University, Australia [2] McCombs School of Business, University of Texas at Austin. Correspondence to: Lan Du <lan.du@monash.edu>, Mingyuan Zhou <mingyuan.zhou@mccombs.utexas.edu>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018. Copyright 2018 by the author(s).

Sample $\boldsymbol{w}_{k_1}^{<s>}$. After integrating $\beta_{vk_1}^{<s>}$ out and ignoring unrelated terms, the joint likelihood related to $\boldsymbol{w}_{k_1}^{<s>}$ is proportional to:

$$\mathcal{L}\left(\pi_{vk_1}^{<s>}\right) \propto \frac{\left(e^{\pi_{vk_1}^{<s>}+\log\log\frac{1}{q_{k_1}}}\right)^{h_{vk_1}^{<s>}}}{\left(1+e^{\pi_{vk_1}^{<s>}+\log\log\frac{1}{q_{k_1}}}\right)^{\alpha_{k_1}^{<s>}+h_{vk_1}^{<s>}}}. \tag{12}$$

The above likelihood can be augmented by an auxiliary variable: $\omega_{vk_1}^{<s>} \sim \mathrm{PG}(1,0)$, where PG denotes the Pólya gamma distribution (Polson et al., 2013). Given $\omega_{vk_1}^{<s>}$, we get:

$$\mathcal{L}\left(\pi_{vk_1}^{<s>}, \omega_{vk_1}^{<s>}\right) \propto e^{\frac{h_{vk_1}^{<s>}-\alpha_{k_1}^{<s>}}{2}\pi_{vk_1}^{<s>}} e^{-\frac{\omega_{vk_1}^{<s>}}{2}\pi_{vk_1}^{<s>2}}. \tag{13}$$

The above likelihood results in the normal likelihood of $\boldsymbol{w}_{k_1}^{<s>}$. Therefore, we sample it from a multi-variate normal distribution as follows:

$$\begin{gathered}
\boldsymbol{w}_{k_1}^{<s>} \sim \mathcal{N}(\boldsymbol{\mu}_{k_1}^{<s>}, \boldsymbol{\Sigma}_{k_1}^{<s>}), \\
\boldsymbol{\mu}_{k_1}^{<s>} = \\
\boldsymbol{\Sigma}_{k_1}^{<s>}\left[\sum_v^V\left(\frac{h_{vk_1}^{<s>}-\alpha_{k_1}^{<s>}}{2} - \omega_{vk_1}^{<s>}\log\log\frac{1}{q_{k_1}}\right)\boldsymbol{f}_v\right], \\
\boldsymbol{\Sigma}_{k_1}^{<s>} = \left[\mathrm{diag}(1/\boldsymbol{\sigma}^{<s>}) + \sum_v^V \omega_{vk_1}^{<s>}\boldsymbol{f}_v(\boldsymbol{f}_v)^\top\right]^{-1},
\end{gathered} \tag{14}$$

where $\boldsymbol{\mu}_{k_1}^{<s>} \in \mathbb{R}^L$ and $\boldsymbol{\Sigma}_{k_1}^{<s>} \in \mathbb{R}^{L\times L}$.

According to (Polson et al., 2013), we can sample

$$\omega_{vk_1}^{<s>} \sim \mathrm{PG}\left(h_{vk_1}^{<s>} + \alpha_{k_1}^{<s>}, \pi_{vk_1}^{<s>} + \log\log\frac{1}{q_{k_1}}\right). \tag{15}$$

To sample from the Pólya gamma distribution, we use an accurate and efficient approximate sampler in Zhou (2016).

Finally, $\boldsymbol{\sigma}^{(s)}$ can be sampled from its gamma posterior.

The inference algorithm is shown in Figure 1.

2. Visualization of the Discovered Topic Hierarchies and Sub-topics

Figure 1-9 show the topic hierarchies discovered by WEDTM on WS, TMN, and Twitter respectively.

3. Generating Synthetic Documents

Below we provide several synthetic documents generated from WEDTM, following the GBN paper (Zhou et al., 2016). Given trained $\{\boldsymbol{\Phi}^{(t)}\}_t$, we used the generative model shown in Figure 1 in the main paper to generate a simulated topic weights $\boldsymbol{\theta}_j^{(1)}$. We show the top words ranked according to $\boldsymbol{\Phi}^{(1)}\boldsymbol{\theta}_j^{(1)}$. Below are some example synthetic documents generated in this manner with WEDTM trained on the TMN dataset. The generated documents are clearly interpretable.

1. study drug cancer risk health people heart women disease patients drugs researchers children brain found finds kids high doctors suggests research medical surgery food diabetes treatment blood years report men young shows year time scientists mets age linked yankees coli long care tuesday good hospital early prostate breast fda developing common adults monday outbreak older weight lower parents problems taking day studies virus aids babies life diet thursday loss levels higher wednesday home evidence make trial tests effective death sox therapy pregnancy experts prevent low patient run pressure pain alzheimer game flu type win number treat start season obesity symptoms

2. mets yankees game season sox win run baseball nfl league hit home time team red inning players day back beat phillies start pitcher innings night runs rangers giants major year games rays big manager york victory indians coach left past boston make draft dodgers hits good pitch hitter career bay tigers blue list play braves fans marlins tuesday years disabled high cubs thursday saturday lockout chicago week nationals top wednesday lead white jays twins streak los days philadelphia field long texas josh pitched ninth friday sunday francisco homer lineup put young college football times roundup pitching cleveland loss sports angeles

3. japan nuclear earthquake plant power tsunami crisis japanese oil quake radiation stocks disaster tokyo prices world friday tuesday fukushima investors hit crippled energy reactor water march monday safety thursday plants economic reactors market government week wednesday year country devastating wall high workers radioactive concerns worst street officials percent companies damage electric report massive fears earnings damaged economy impact food levels agency operator global daiichi plans markets stock chernobyl sales coast growth rise higher states atomic fuel united risk stricken health crude supply dollar low strong demand level recovery day magnitude quarter shares fell struck tepco billion commodities experts gains relief

4. wedding royal prince kate william middleton london queen britain ireland british irish friday eliza-

```
Require: $\{\boldsymbol{x}_j^{(1)}\}_j, T, S, \{K_t\}_t, a_0, b_0, e_0, f_0, \eta_0, MaxIteration$
Ensure: The sampled value for all the latent variables

Randomly initialise all the latent variables according to the generative process;
for $t \leftarrow 1$ to $T$ do
  for $iter \leftarrow 1$ to $MaxIteration$ do
    if $t = 1$ then
      $/ *$ Generating documents $* /$
      Sample the bottom-layer topics by Eq. (1); Calculate $\{x_{vjk_1}^{(1)}\}_{v,j,k_1}$;
    end if
    $/ *$ Inter topic structure $* /$
    Do the upward-downward Gibbs sampler
    (Algorithm 1 in Zhou et al. (2016)) for $\{\boldsymbol{\theta}_j^{(t)}, \boldsymbol{c}_j^{(t)}, \boldsymbol{p}_j^{(t)}\}_{t,j}, \{\boldsymbol{\Phi}^{(t)}\}_t, \boldsymbol{r}$;
    if $t = 1$ then
      $/ *$ Intra topic structure $* /$
      Sample $\{q_{k_1}\}_{k_1}$ by Eq. (3); Sample $\{h_{vk_1}\}_{v,k_1}$ by Eq. (4);
      for $s \leftarrow 1$ to $S$ do
        Sample $\{h_{vk_1}^{<s>}\}_{v,k_1}$ by Eq. (6);
        Sample $\{g_{vk_1}^{<s>}\}_{v,k_1}$ by Eq. (11); Calculate $\{g_{\cdot k_1}^{<s>}\}_{k_1}$ by Eq. (10);
        Sample $\alpha_0^{<s>}$ and $c_0^{<s>}$ from their gamma posterior;
        Sample $\{\alpha_{k_1}^{<s>}\}_{k_1}$ by Eq. (9);
        Sample $\{\omega_{vk_1}^{<s>}\}_{v,k_1}$ by Eq. (15); Sample $\boldsymbol{\sigma}^{<s>}$ from its gamma posterior;
        Sample $\{\boldsymbol{w}_{k_1}^{<s>}\}_{k_1}$ by Eq. (14);
        Calculate $\{\pi_{vk_1}^{<s>}\}_{v,k_1}$ by Eq. (8);
        Sample $\{\beta_{vk_1}^{<s>}\}_{v,k_1}$ by Eq. (7);
      end for
      Calculate $\{\beta_{vk_1}\}_{v,k_1}$;
    end if
  end for
end for
```

Figure 1. Gibbs sampling algorithm for WEDTM

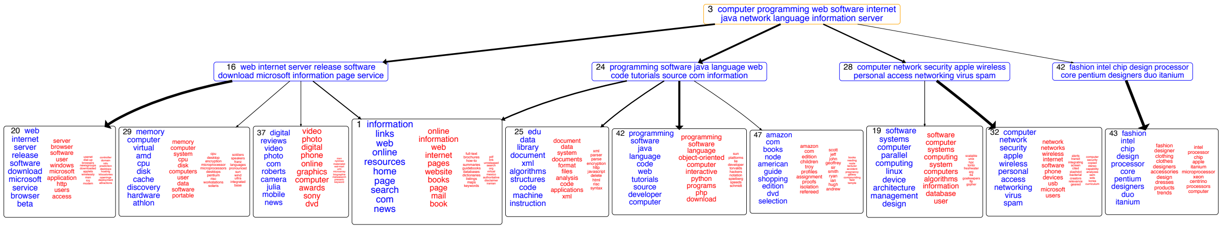
3 computer programming web software internet java network language information server
16 web internet server release software download microsoft information page service
24 programming software java language web code tutorials source com information
28 computer network security apple wireless personal access networking virus spam
42 fashion intel chip design processor core pentium designers duo itanium
20 web internet server release software download microsoft service browser beta
server browser software user windows microsoft application http users access
29 memory computer virtual amd cpu disk cache discovery hardware athlon
memory computer system cpu disk computers user data software portable
37 digital reviews video photo com roberts camera julia mobile news
video photo digital phone online graphics computer awards sony dvd
1 information links web online resources home page search com news
online information web internet pages website books page mail book
25 edu data library document xml algorithms structures code machine instruction
document data system documents format files analysis code applications xml
42 programming software java language code web tutorials source developer computer
programming software language object-oriented computer interactive python programs php download
47 amazon com books node american guide shopping edition dvd selection
amazon com edition children troy profiles assignment proofs isolation refereed
19 software systems computer parallel computing linux device architecture management design
software computer systems computing system computers algorithms information database user
32 computer network security apple wireless personal access networking virus spam
network networks wireless internet software phone devices usb microsoft users
43 fashion intel chip design processor core pentium designers duo itanium
fashion designer clothing clothes designers accessories design dresses products trends
intel processor chip apple itanium microprocessor xeon centrino processors computer

Figure 2. Analogous plots to Figure 6 in the main paper for a tree about “computers” on the WS dataset.

beth designer princess april couple week palace people visit idol american day fashion world made abbey art bride marriage diana duchess secret time honeymoon photo king dress dinner famous sarah cake back media lady buckingham year ring show photos hat coverage wednesday party saturday mcqueen month morning pop watch guests charlie days sheen westminster future kardashian star television night work crown tribute ago duke officials worn check years pre kim maya wear final site met ceremony music tonight museum nuptials harry ferguson pounds royals trip tuesday turned

5. apple sony mobile data ipad company corp network phone billion software google iphone computer mil-

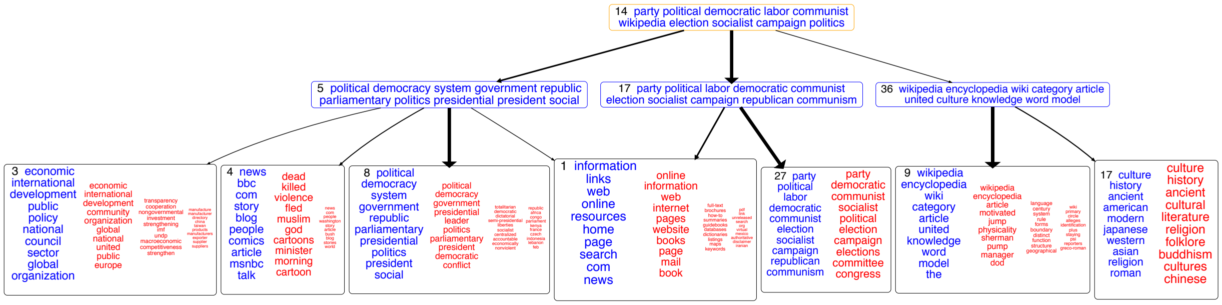
14 party political democratic labor communist wikipedia election socialist campaign politics
5 political democracy system government republic parliamentary politics presidential president social
17 party political labor democratic communist election socialist campaign republican communism
36 wikipedia encyclopedia wiki category article united culture knowledge word model
3 economic international development public policy national council sector global organization
4 news bbc com story blog people comics article msnbc talk
8 political democracy system government republic parliamentary presidential politics president social
1 information links web online resources home page search com news
27 party political labor democratic communist election socialist campaign republican communism
9 wikipedia encyclopedia wiki category article united knowledge word model the
17 culture history ancient american modern japanese western asian religion roman

Figure 3. Analogous plots to Figure 6 in the main paper for a tree about “movie” on the WS dataset.

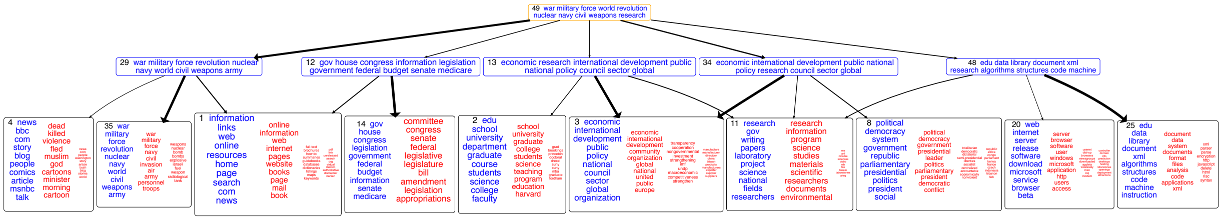
49 war military force world revolution nuclear navy civil weapons research
29 war military force revolution nuclear navy world civil weapons army
12 gov house congress information legislation government federal budget senate medicare
13 economic research international development public national policy council sector global
34 economic international development public national policy research council sector global
48 edu data library document xml research algorithms structures code machine
4 news bbc com story blog people comics article msnbc talk
35 war military force revolution nuclear navy world civil weapons army
1 information links web online resources home page search com news
14 gov house congress legislation government federal budget information senate medicare
2 edu school university department graduate course students science college faculty
3 economic international development public policy national council sector global organization
11 research gov writing papers laboratory project science national fields researchers
8 political democracy system government republic parliamentary presidential politics president social
20 web internet server release software download microsoft service browser beta
25 edu data library document xml algorithms structures code machine instruction

Figure 4. Analogous plots to Figure 6 in the main paper for a tree about “war” on the WS dataset.

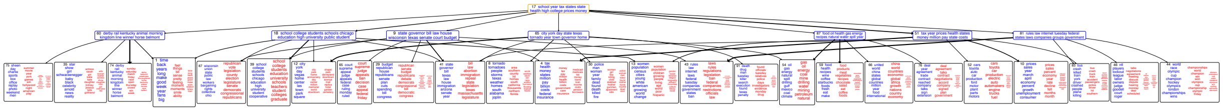

Figure 5. Analogous plots to Figure 6 in the main paper for a tree about “daily life” on the TMN dataset.

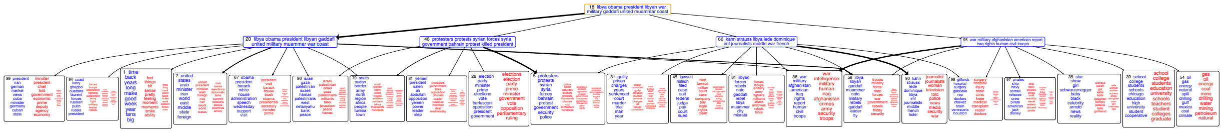
18 libya obama president libyan war military gaddafi united muammar coast

Figure 6. Analogous plots to Figure 6 in the main paper for a tree about “politics” on the TMN dataset.

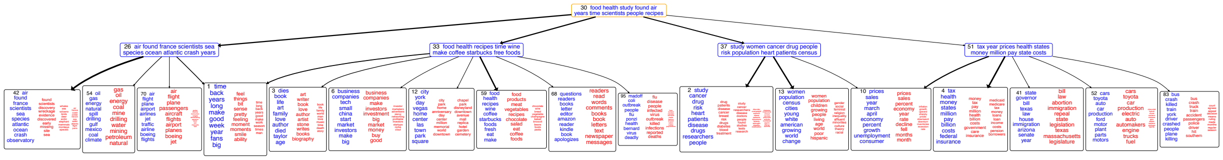
30 food health study found air years time scientists people recipes
26 air found france scientists sea species ocean atlantic crash years
33 food health recipes time wine make coffee starbucks free foods
37 study women cancer drug people risk population heart patients census
51 tax year prices health states money million pay state costs

Figure 7. Analogous plots to Figure 6 in the main paper for a tree about “food and health” on the TMN dataset.

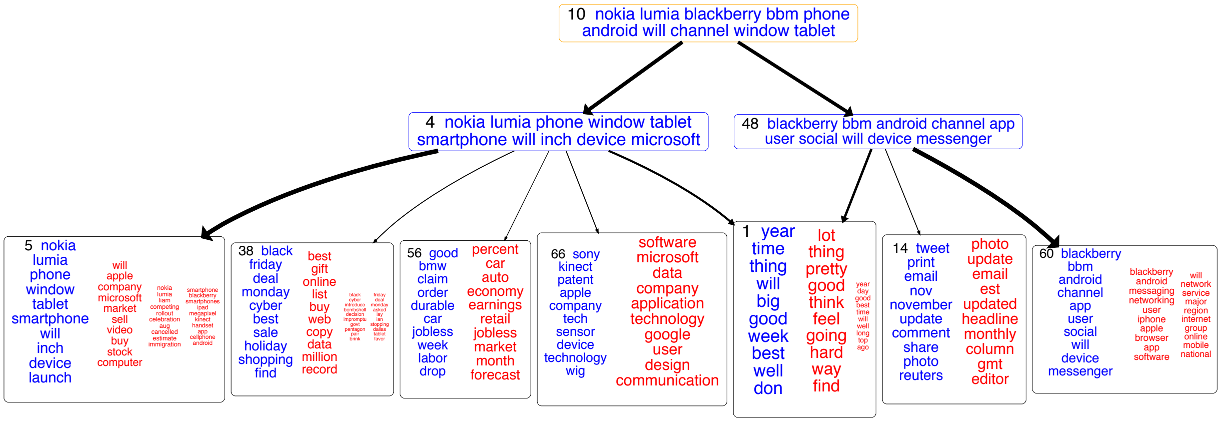
10 nokia lumia blackberry bbm phone android will channel window tablet
4 nokia lumia phone window tablet smartphone will inch device microsoft
48 blackberry bbm android channel app user social will device messenger
5 nokia lumia phone window tablet smartphone will inch device launch
38 black friday deal monday cyber holiday shopping find
56 good bmw claim order durable car jobless week labor drop
66 sony kinect patent apple company tech sensor device technology wig
1 year time thing will big good week best well don
14 tweet print email nov november update comment share photo reuters
60 blackberry bbm android channel app user social will device messenger

Figure 8. Analogous plots to Figure 6 in the main paper for a tree about “phone” on the Twitter dataset.

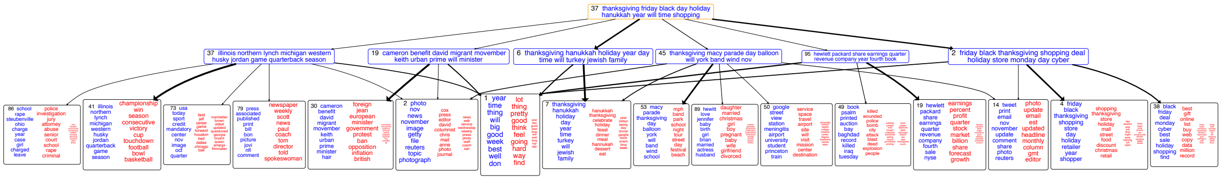
37 thanksgiving friday black day holiday hanukkah year will time shopping
37 illinois northern lynch michigan western husky jordan game quarterback season
19 cameron benefit david migrant movember keith urban prime will minister
6 thanksgiving hanukkah holiday year day time will turkey jewish family
45 thanksgiving macy parade day balloon will york band wind nov
95 hewlett packard share earnings quarter revenue company year fourth book
2 friday black thanksgiving shopping deal holiday store monday day cyber

Figure 9. Analogous plots to Figure 6 in the main paper for a tree about “thanksgiving” on the Twitter dataset.

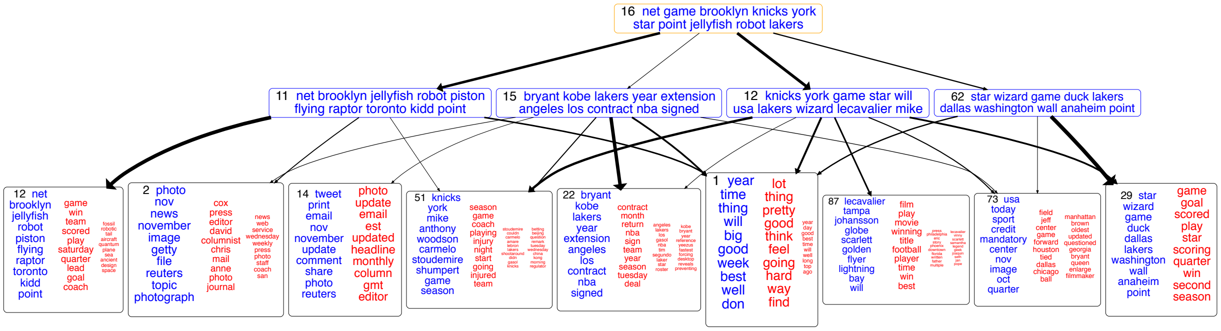
16 net game brooklyn knicks york star point jellyfish robot lakers
11 net brooklyn jellyfish robot piston flying raptor toronto kidd point
15 bryant kobe lakers year extension angeles los contract nba signed
12 knicks york game star will usa lakers wizard lecavalier mike
62 star wizard game duck lakers dallas washington wall anaheim point
12 net brooklyn jellyfish robot piston flying raptor toronto kidd point
2 photo nov news november image getty file reuters topic photograph
14 tweet print email nov november update comment share photo reuters
51 knicks york mike anthony woodson carmelo stoudemire shumpert game season
22 bryant kobe lakers year extension angeles los contract nba signed
1 year time thing will big good week best well don
87 lecavalier tampa johansson globe scarlett golden flyer lightning bay will
73 usa today sport credit mandatory center nov image oct quarter
29 star wizard game duck dallas lakers washington wall anaheim point

Figure 10. Analogous plots to Figure 6 in the main paper for a tree about “NBA” on the Twitter dataset.

lion microsoft market amazon tablet playstation maker devices wireless app technology security service sales users year smartphone android customers phones deal plans nokia report world week system online breach smartphones services amp apps group percent buy executive top thursday music information ceo computers samsung consumer location chief wednesday stores largest research business electronics growth books prices jobs china windows major selling hackers firm launch high digital tuesday friday shares intel companies systems tablets make years device city people population energy verizon blackberry profit internet bid store

Below are some example synthetic documents with WEDTM trained on the WS dataset.

1. research papers thesis project writing edu paper dissertation proposal reports patent technical report topics guide academic exploratory information write methods researchers publications parallel process phd issues ideas computing media topic custom projects essay help archives tools page labs web lab university communication resources ibm proposals intellectual home advice cluster dissertations survey library essays address practical links step abstract property bell systems linux performance master serve abstracts developed written department index quality exploring matters res theses documents course college collaborative processing programme materials focused preparing computer program students search guidelines online articles experience services recommendations welcome conduct funding aspects idea people

2. football team league news american fans nfl players soccer stadium sports indoor national com home statistics club history hockey baseball college fan association teams arena england game player texas website scores basketball professional fixtures espn season ncaa clubs university afl schedule tickets minnesota independent tables athletics sport welcome united leagues dedicated ice chicago giants nba stadiums online rugby stats nhl gymnastics coverage coaches levels delivers city arsenal index coach database conference standings information cstv fantasy bears articles views athletic moved rules assault bulls supporters officials kickoff division games women tournament photos body media fame covering adjacent sexual federation pages rutgers

3. health hiv prevention stress healthy aids diet nutrition hepatitis fitness food epidemic pressure infection information cdc disease people blood life gov diseases epidemiology treatment cholesterol picnic tips medical articles children safety symptoms fat topics heart foods living centers fatty help kids liver care researchers eating diagnosis who weight body virus recipes human control viral sheets com helps patterns and comprehensive diabetes risk trials int learn conditions advice influenza diets exercise affect public press consensus constant united eat infections listing unaids experts mind personality disorder causes world flu infectious guidelines global loss women avian humans population ucsf cope job brains index

4. system government parliamentary republic presidential maradona diego parliament systems westminster freedom independence democracy semi-presidential consensus constitutional armando elected congressional constitution executive america people declaration elections president czech gov australia election documents hand french political national congress authority united archives legislature kingdom country governments powers power public british representation canada argentine versus buenos separation born legislative central assembly affairs republics india peace france sri semi english branch body operates charters romania distributed presidency practice the parliaments governing house countries britain rule federal principles jury lanka foreign bill govern nunavut portuguese ireland aires immense head qld parl pub shtml politics ukraine parties

5. school research college edu graduate university education students elementary center student library district information schools program programs degree city public media papers phd master thesis law department undergraduate home project staff america regional writing scholarship calendar independent academic town paper administration news scholarships north dissertation community faculty welcome archives virginia children president union wisconsin temple proposal parents ohio pennsylvania teachers philadelphia reports guide houston nation degrees chronicle services massachusetts patent opportunities boston technical activities county study masters east report employment commission usa expo lewis michigan san academy memorial topics meeting institute campus homepage web arizona maine business announcements philly hub

References

Polson, N., Scott, J., and Windle, J. Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, 108(504): 1339–1349, 2013.

Wray, B. and Marcus, H. A Bayesian view of the Poisson-Dirichlet process. *arXiv preprint arXiv:1007.0296v2 [math.ST]*, 2012.

Zhao, H., Rai, P., Du, L., and Buntine, W. Bayesian multi-label learning with sparse features and labels, and label co-occurrences. In *AISTATS*, pp. 1943–1951, 2018.

Zhou, M. Softplus regressions and convex polytopes. *arXiv preprint arXiv:1608.06383*, 2016.

Zhou, M. and Carin, L. Negative binomial process count and mixture modeling. *TPAMI*, 37(2):307–320, 2015.

Zhou, M., Cong, Y., and Chen, B. Augmentable gamma belief networks. *JMLR*, 17(163):1–44, 2016.