Supplementary Materials for "Topic-Partitioned Multinetwork Embeddings"

Peter Krafft*
CSAIL
MIT
pkrafft@mit.edu

Juston Moore†, Bruce Desmarais‡, Hanna Wallach†
†Department of Computer Science, ‡Department of Political Science
University of Massachusetts Amherst
†{jmoore, wallach}@cs.umass.edu
‡desmarais@polsci.umass.edu

1 Generative Process

The complete generative process for our model is as follows:

for topic $t = 1, \ldots, T$ **do**
  Topic-specific distribution over word types $\boldsymbol{\phi}^{(t)} \sim \mathrm{Dir}(\beta)$
  for actor $a = 1, \ldots, A$ **do**
    Topic-specific point $\boldsymbol{s}_a^{(t)} \sim \mathcal{N}(\mathbf{0}, \sigma_1^2 \boldsymbol{I})$
  end for
  Topic-specific bias $b^{(t)} \sim \mathcal{N}(\mu, \sigma_2^2)$
end for
for email $d = 1, \ldots, D$ **do**
  Email-specific distribution over topics $\boldsymbol{\theta}^{(d)} \sim \mathrm{Dir}(\alpha)$
  for position $n = 1, \ldots, \max(1, N^{(d)})$ **do**
    Topic assignment $z_n^{(d)} \sim \boldsymbol{\theta}^{(d)}$
    if $N^{(d)} \neq 0$ **then**
      Token $w_n^{(d)} \sim \boldsymbol{\phi}^{(z_n^{(d)})}$
    end if
  end for
  for recipient $r = 1, \ldots, A$ **do**
    if $r \neq a^{(d)}$ **then**
      Position assignment $x_r^{(d)} \sim \mathrm{U}(1, \ldots, \max(1, N^{(d)}))$
      Recipient indicator $y_r^{(d)} \sim \mathrm{Bern}(p_{a^{(d)}r}^{(t)})$ assuming $z_{x_r^{(d)}}^{(d)} = t$
    end if
  end for
end for

The corresponding directed graphical model is shown in Figure 1.

2 Definitions of Discrepancy Functions

The dyad intensity distribution [1] quantifies the level of node–node activity in a network. Letting $N^{(1|a,r)} = \sum_{d=1}^{D} \mathbf{1}(a^{(d)} = a)\, \mathbf{1}(y_r^{(d)} = 1)$,[1] the dyad intensity for actors a and r is the geometric

*Work done at the University of Massachusetts Amherst
[1]The function $\mathbf{1}(\cdot)$ evaluates to one if its argument evaluates to true and evaluates to zero otherwise.

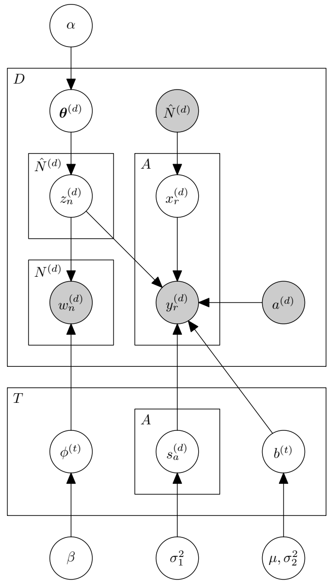

Figure 1: The directed graphical model for our model. Count $\hat{N}^{(d)}$ is equal to $\max(N^{(d)}, 1)$, indicating that there is a single, "dummy" topic assignment $z_1^{(d)}$ for any email that contains no text.

mean of the number of emails sent from a to r and the number of emails sent from r to a:

$$f_1(a, r; \mathcal{Y}) = \left(N^{(1|a,r)}\, N^{(1|r,a)}\right)^{\frac{1}{2}}.$$

The vertex degree of actor a is the number of emails sent to or from a:

$$f_2(a; \mathcal{Y}) = \sum_{r=1}^{A} \left(N^{(1|a,r)} + N^{(1|r,a)}\right).$$

The vertex degree distribution is the distribution of vertex degrees in a network.

The geodesic distance distribution [2] quantifies the connectivity of a network. We define the pairwise distance from actor a to actor r to be $1 \,/\, N^{(1|a,r)}$. The geodesic distance from any actor a' to any actor r' is the shortest path between them in the weighted graph induced by the set of pairwise distances. The geodesic distance distribution is the distribution of geodesic distances.

Generalized graph transitivity [3] quantifies the amount of clustering in a weighted network. We define a triple to be any set of three actors a, r, and r', such that there is at least one email from a to r and from r to r'. If there are also one or more emails from a to r', then triple (a, r, r') is a transitive triple. Letting the value of triple (a, r, r') be $N^{(1|a,r)} + N^{(1|r,r')}$, generalized graph transitivity is defined as the ratio of the sum of the transitive triples' values to the sum of all triples' values.

Acknowledgments

This work was supported in part by the Center for Intelligent Information Retrieval and in part by the NSF GRFP under grant #1122374. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsors.

References

[1] D. Wyatt, T. Choudhury, and J. Bilmes. Discovering long range properties of social networks with multi-valued time-inhomogeneous models. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[2] J.H. Fowler. Legislative cosponsorship networks in the us house and senate. *Social Networks*, 28(4):454–465, 2006.

[3] T. Opsahl and P. Panzarasa. Clustering in weighted networks. *Social Networks*, 31(2):155–163, 2009.