

Topic Modeling on Health Journals with Regularized Variational Inference

Challenge. CaringBridge data: 14M journals written by 500k patients and caregivers during a health crisis.

- Many authors writing asynchronously over time.

Approach. Design new Dynamic Author-Persona (DAP) topic model.

- Represent authors by a persona — personas capture propensity to discuss certain topics over time.

Novelty. Regularized Variational Inference: Encourages certain solutions.

- Nudge DAP's personas to be distinct.

State-of-the-Art Performance + Compelling Narratives

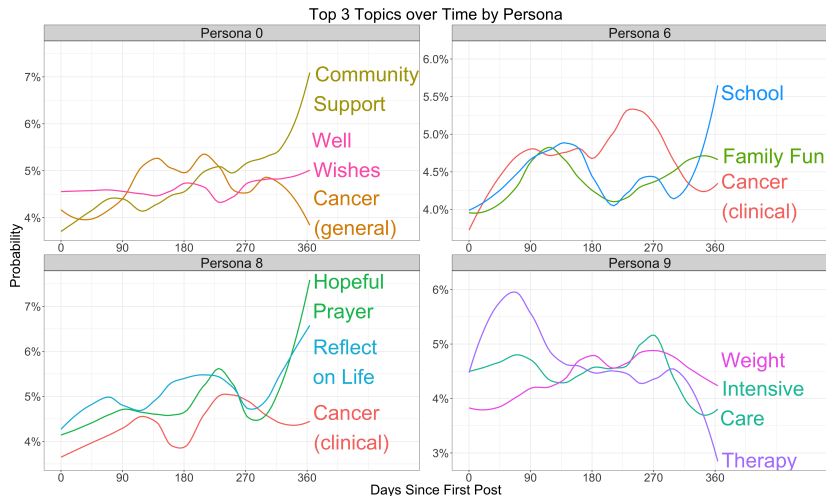


Figure: DAP finds personas that reflect common "health journeys" shared by CaringBridge users. Each plot shows how the 3 topics most associated with a persona evolve over time. Topic labels are hand-selected based on top words in each topic.