Shared Topics in Social Networks

Laura Dietz dietz@informatik.uni-mannheim.de
Social Network With Contents

- friendship / citation
- user / paper
- items / content
APPLICATIONS
Application: Filter Citation Graphs

- Seed publications.
- Add citation vicinity.
- Add cited publications.

- Infer citation strengths.
- Filter edges.
- Filter isolated nodes.
Application: Filter Citation Graphs

- Seed publications.
- Add citation vicinity.
- Add cited publications.

- Infer citation strengths.
- Filter edges.
- Filter isolated nodes.
Taste-based Subscription

- Subscribe to contents of a friend.
- Infer the common taste.
- Filter / re-weight the friend’s contents.

Me ➔ Friend of mine

Me ➔ Yuck!

Me ➔ Yeah!
Applications

- Filter citation graphs
- Taste-based subscription
- Visualize the neighborhood: colors / sort / filter
- Recommend new friends / items / user groups
NETWORK TOPIC MODELS
Outline

- Latent Dirichlet Allocation
- Links-as-Vocabulary-Model
- Stochastic Blockmodel
- Citation Influence Model
- Shared Taste Model
- Results & Comparison
Latent Dirichlet Allocation (Forward)

for each cited doc...
for each token...
for each topic...
  topic for
word in
cited doc
w'

for each cited doc...
for each topic...
  word distr
for each topic...
Latent Dirichlet Allocation (Inference)

- Which Topic?
- Other topics in same doc.
- Topics of same words in any doc.

- dirichlet prior \( \alpha_\theta \)
- topic mixture of cited pub \( \theta \)
- topic for word in cited doc \( t' \)
- word in cited doc \( w' \)
- dirichlet prior \( \alpha_\phi \)
- word distr for each topic \( \phi \)

Typical words for each topic

for each token...
for each topic...
for each cited doc...
for each token...
for each cited doc...
How to Factor in the Network?
Links as Vocabulary

- Links as Vocabulary (Cohn&Hofman, Erosheva,..)
  - Topics have typical words and typical links

- Issue: weak coupling between content and links.
- Issue: Links may be outnumbered by words.
Blockmodel-Style

- Blockmodel-Style (Nallapati, Chang&Blei, Erosheva)
  - Only linked nodes should have similar topics.

  Issue: Links are absent due to several reasons.
Shared Topics and Influence of Linked Vertices

Prediction of Citation Influences
(Dietz, Bickel, Scheffer, 2007 ICML)
Plain Citation Influence Model (Forward)
Plain Citation Influence Model (Inference)
Topic Coupling

- $\theta$ is about $c$, and all citing pubs $d_1,d_2,…$ (partially)
- Coupling between:
  - Citing and cited publications.
  - Bibliographically coupled publications.
Evolving Terminology

- Evolving terminology
  - i.e., latent Dirichlet allocation <-> probabilistic topic model
  - Indirectly influences the “copy process” for other nodes.
Citation Influence Model With Own Topics

- **Issue**: all words have to be associated to a citation.
  - Introduces noise in the shared topic mixture.
  - Influence association process of other citing publications.

- **Solution**: Model may decide leave some words unassociated.
Bipartite Transformation

- Transform citation graph to bipartite graph:
  - Duplication of nodes.
  - Don’t worry! Topics in clones will be similar (because of co-occurrence patterns)

- Allows for application to general graphs!
Experiments: Predictive Performance

- Citation influence best method on average.
  - Significantly better than LDA (paired-t-test, $\alpha=5\%$).
- Robust towards the number of topics.
- TF-IDF and PageRank do not work (AUC$\approx0.5$).
Shared Topics and Influence of Linked Vertices

Shared Tastes in Online Communities

(Dietz, Gamari, Guiver, Snelson, and Herbrich, 2012, ICWSM)
Shared Taste Model (Forward)

For all items of user $u$

For all users $u$

For all friendships $(u,f)$

For all friends $f$

Discrete

Dirichlet

taste

distr

friend

prior

item

prior

item distr

Tastes
Shared Taste Model (Forward)

For all items of user $u$

For all items of user $u$

For all friendships $\{u,f\}$

For all users $u$

For all friends $f$
Shared Taste Model (Inference)

- Matching items of both friends are assigned to the edge.
- Matching items => edge’s topic mixture.
- The more items are assigned to a particular edge the higher the influence of the opposite node.
- Typical sharing patterns => item distribution per topic.
Shared Taste vs.

- No node duplication
- Topics: One friendship
- Frienddistr = Compatibility

Citation Influence

- Nodes as friends / users
- Topics: Me & my friends
- Influence = Compatibility
Jurassic Park
by Michael Crichton

Series: Jurassic Park (1)

Members | Reviews | Popularity | Average rating | Conversations
---------|---------|------------|----------------|-----------------
7,932    | 88      | 218        | ★★★★ (3.92)   | 81

Recently added by: wampyrii, Gryphor, kmmeinert, Smcewen, carlogulin, Comabunny, svfreeman, scmerritt, LEGIOXIII, Echo049

Tags:
- 20th century
- action
- adventure
- american
- chaos theory
- cloning
- crichton
- dinosaurs
- dna
- fantasy
- fiction
- genetic engineering
- genetics
- horror
- jurassic park
- made into movie
- michael crichton
- movie
- novel
- own paperback
- read
- sci-fi
- science
- science fiction
- sf
- suspense
- technothriller
- thriller
Results on Librarything

![Graph showing ROC-AUC for different model derivatives. The x-axis represents various model derivatives (Base, ST, SST, CI, B), and the y-axis represents the ROC-AUC values ranging from 0.5 to 0.7. Different markers represent different model variations: Shared taste (ST), Shared taste own topics (STt), Shared taste own items (STi), Latent Dirichlet allocation (LDA), Popularity (pBase), Citation influence (CI), Citation influence own topics (CIt), Citation influence own items (CII), Trivial (tBase), Genre (gBase), Single shared taste (SST), Single shared taste own topics (STt), Single shared taste own items (STi), Symmetric blockmodel (sB), Asymmetric blockmodel (aB).]
Shared taste model with own topics

c) Group-topic correlation

d) Tag clouds according to $\phi_t$

t6: Computer & Scifi

t4: History & Politics

t7: Scifi & Theory

t1: Fantasy & Scifi

Citation influence model (with own topics)

e) Topic-group correlation

f) Tag clouds according to $\phi_t$

t3: computer & sci-fi 1

t10: computer & sci-fi 2

t4: Female’s readings 1

t6: Female’s readings 2
When to use which?

- Which assumptions hold for your data set?
  - Coherent communities: Citation Influence Model
  - Overlapping facets: Shared Taste Model
  - Absence of links important: Blockmodel

- What do you want to know?
  - Influence strength: Cit Inf / Shared Taste
  - Individual Topics: Blockmodel
  - Topics of two friends: Shared Taste
  - Topics of neighborhoods: Cit Inf / Links-as-Vocabulary
Summary: Shared Topics in Social Networks

- Exploiting network structure in topic models.
- Code online: http://github.com/bgamari/bayes-stack

- Reusing topics = compatible nodes = strong influence.
- Citation influence model
  - Using other node’s topics.
- Shared taste model
  - Using topics from a common edge.