Managing Online Business Communities

The ROBUST Project
Steffen Staab, Thomas Gottron
& ROBUST Project Team
Business Communities

- Information ecosystems
  - Employees
  - Business Partners, Customers
  - General Public

Valuable asset

Opportunities

Risks
Use Cases

Lotus Connections
Communities
- Employees
- Working groups
- Interest Groups
- Projects

Business value
- Task relevant information
- Collaboration
- Innovation

Volume
- 4,000 posts/day
- 386,000 employees
- 1.5GB content/day

IBM

SAP Community Network (SCN)
Communities
- Customers
- Partners
- Suppliers
- Developers

Business value
- Products support
- Services
- Find business partners

Volume
- 6,000 posts/day
- 1,700,000 subscribers
- 16GB log/day

SAP

MeaningMine
Communities
- Social media
- News
- Web fora
- Public communities

Business value
- Topics
- Opinions
- Service for partners

Volume
- 1,400,000 posts/day
- 708,000 web sources
- 45GB content/day

Polecat

Employees
Intranet

Business Partners
Extranet

Public Domain
Internet
High Level Architecture

- Risk Dashboard
- Community Data
- Cloud Based Data Management
- Community Simulation
- Community Risk Management
- User Roles and Models
- Community Analysis
- Communication Bus

Community Data

ROBUST

WeST
People and Knowledge Networks
Requirements

Cloud Based Data Management
- Operations on Data
- Scalability
- Real Time
- Parallel Execution
- Stream Based

Community Risk Management
- Risk Formalization
- Risk Detection
- Risk Forecasting
- Risk Management
- Risk Visualization

Community Analysis
- Feature Selection
- Structural Analysis
- Behavior Analysis
- Content Analysis
- Cross Community

Community Simulation
- Community Models
- Policy Models
- Influence
- Simulation
- Prediction

User Roles and Models
- User Needs
- Motivation
- Roles
- Groups
- Community Value
The ROBUST platform

Risk Dashboard and Visualizations
A new Role: the Community Manager

- Definition of risks/opportunities
- Monitoring
- Interaction with community
- Reaction and countermeasures

→ Need for a control center
Risk matrix visualization

Risks ordered by probability

Risk/Opportunity matrix

- R1: Opportunity to gain control
- R2: Risk of undesirable role change
- R3: Risk of community backlash
- R4: Risk of key contributors leaving

A single risk can have multiple impacts.

Positive impacts

Negative impacts
Community Analysis,
User Roles & Risk Management

Determine Risk:
Users Leaving the Community
“Churn”

• Users churn = Users leave the community, become inactive

• Questions:
  – Why do users leave the community?
  – Who is leaving the community?
  – Impact?

• Tasks:
  – Detect “Churn”
  – Predict “Churn”
  – Evaluate “Churn”
What is Churn? Activity?

Activity bursts
No churn

Holiday
No churn

Inconsistent
No churn

Cease to be active
Churn!
Why Churn?
Network Effects

- Churn: churning users influence other users they communicate with

- Temporal ordering: Churners that churned subsequent to each other

- Frequently, after a slow start, resulting in a cascade of churning
Role Analysis: who churns?

- A role represents the standing, or part, that a user has within a given community

<table>
<thead>
<tr>
<th>Role \ Dimension</th>
<th>Reciprocity</th>
<th>Initialisation</th>
<th>Persistence</th>
<th>Popularity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elitist</td>
<td>high: Threads</td>
<td></td>
<td></td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>low: Neighbours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joining Conversationalist</td>
<td>low</td>
<td>low high</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Popular Initiator</td>
<td>high</td>
<td></td>
<td>high</td>
<td></td>
</tr>
<tr>
<td>Supporter</td>
<td>medium</td>
<td>low</td>
<td></td>
<td>medium</td>
</tr>
<tr>
<td>Taciturn</td>
<td>low</td>
<td></td>
<td>low</td>
<td></td>
</tr>
<tr>
<td>Ignored</td>
<td></td>
<td></td>
<td></td>
<td>low: No replies</td>
</tr>
</tbody>
</table>
Reciprocity, initialisation, persistence, popularity

Feature levels change with the dynamics of the community

Based on related work, we associate roles with a collection of feature-to-level mappings e.g. in-degree -> high, out-degree -> high

Run rules over each user’s features and derive the community role composition
Churn Analysis on Roles

- **Role composition in community**

  ![Pie chart for Week 1]

  - Grunts: 16%
  - Pop. Partic.: 3%
  - Pop. Init.: 11%
  - Support.: 40%
  - Ignored: 12%

- **Development towards an unhealthy role composition!**
Community & Content Analysis

Opportunity Detection: Discovering Interesting Contents
Analogies to Twitter!
RT @janedoe: My dear @johndoe had troubles to wake up this #morning
Retweets

- Retweet indicates quality
  - „of interest for others“

- Idea:
  - Learn to predict retweets!

Likelihood of retweet as metric for Interestingness

Naveed et al. WebSci 2011
Aim: Prediction of probability of retweet

Logistic regression:

\[ f(z) = \frac{1}{1+e^{-z}} \]

\[ z = w_0 + w_1 x_1 + \cdots + w_n x_n \]

Model parameters \( w_i \) learned on training data

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Users</th>
<th>Tweets</th>
<th>Retweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choudhury</td>
<td>118,506</td>
<td>9,998,756</td>
<td>7.89%</td>
</tr>
<tr>
<td>Choudhury (extended)</td>
<td>277,666</td>
<td>29,000,000</td>
<td>8.64%</td>
</tr>
<tr>
<td>Petrovic</td>
<td>4,050,944</td>
<td>21,477,484</td>
<td>8.46%</td>
</tr>
<tr>
<td>Feature</td>
<td>Dimensions</td>
<td>Weight</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>(intercept)</td>
<td>-5.45</td>
<td></td>
</tr>
<tr>
<td>Message feature</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct message</td>
<td></td>
<td>-147.89</td>
<td></td>
</tr>
<tr>
<td>Username</td>
<td></td>
<td>146.82</td>
<td></td>
</tr>
<tr>
<td>Hashtag</td>
<td></td>
<td>42.27</td>
<td></td>
</tr>
<tr>
<td>URL</td>
<td></td>
<td>249.09</td>
<td></td>
</tr>
<tr>
<td>Sentiment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence</td>
<td></td>
<td>-26.88</td>
<td></td>
</tr>
<tr>
<td>Arousal</td>
<td></td>
<td>33.97</td>
<td></td>
</tr>
<tr>
<td>Dominance</td>
<td></td>
<td>19.56</td>
<td></td>
</tr>
<tr>
<td>Emoticons</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td>-21.8</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td>9.94</td>
<td></td>
</tr>
<tr>
<td>Exclamation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td>13.66</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td>8.72</td>
<td></td>
</tr>
<tr>
<td>Punctuation</td>
<td>!</td>
<td>-16.85</td>
<td></td>
</tr>
<tr>
<td>?</td>
<td></td>
<td>23.67</td>
<td></td>
</tr>
<tr>
<td>Terms</td>
<td>Odds</td>
<td>19.79</td>
<td></td>
</tr>
</tbody>
</table>
## Logistic Regression: Topic Weights

<table>
<thead>
<tr>
<th>Topic</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>social media market post site web tool traffic network</td>
<td>27.54</td>
</tr>
<tr>
<td>follow thank twitter welcome hello check nice cool people</td>
<td>16.08</td>
</tr>
<tr>
<td>credit money market business rate economy home</td>
<td>15.25</td>
</tr>
<tr>
<td>christmas shop tree xmas present today wrap finish</td>
<td>2.87</td>
</tr>
<tr>
<td>home work hour long wait airport week flight head</td>
<td>-14.43</td>
</tr>
<tr>
<td>twitter update facebook account page set squidoo check</td>
<td>-14.43</td>
</tr>
<tr>
<td>cold snow warm today degree weather winter morning</td>
<td>-26.56</td>
</tr>
<tr>
<td>night sleep work morning time bed feel tired home</td>
<td>-75.19</td>
</tr>
</tbody>
</table>
Re-Ranking using Interestingness

- Top-k interesting tweets for „beer“

<table>
<thead>
<tr>
<th>Rang</th>
<th>Username</th>
<th>Tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BeeracrossTX</td>
<td>UK beer mag declares &quot;the end of beer writing.&quot; @StanHieronymus says not so in the US. <a href="http://bit.ly/424HRQ">http://bit.ly/424HRQ</a> #beer</td>
</tr>
<tr>
<td>2</td>
<td>narmmusic</td>
<td>beer summit @bspward @jhinderaker no one had billy beer? heehee #narm - beer summit @bspward @jhinde <a href="http://tinyurl.com/n29oxj">http://tinyurl.com/n29oxj</a></td>
</tr>
<tr>
<td>3</td>
<td>beeriety</td>
<td>Go green and turn those empty beer bottles into recycled beer glasses!</td>
</tr>
<tr>
<td>4</td>
<td>hblackmon</td>
<td>Great Divide beer dinner @ Porter Beer Bar on 8/19 - $45 for 3 courses + beer pairings. <a href="http://trunc.it/172wt">http://trunc.it/172wt</a></td>
</tr>
<tr>
<td>5</td>
<td>nycraftbeer</td>
<td>Interesting Concept-Beer Petitions.com launches&amp;hopes 2help craft beer drinkers enjoy beer they want @their fave pubs. <a href="http://bit.ly/11gJQN">http://bit.ly/11gJQN</a></td>
</tr>
<tr>
<td>6</td>
<td>carichardson</td>
<td>Beer Cheddar Soup: Dish number two in my famed beer dinner series is Beer Cheddar Soup. I hadn’t had too.. <a href="http://bit.ly/1diDdF">http://bit.ly/1diDdF</a></td>
</tr>
<tr>
<td>8</td>
<td>delphiforum</td>
<td>Love beer? Our member is trying to build up a new beer drinker's forum. Grab a #beer and join us: <a href="http://tr.im/pD1n">http://tr.im/pD1n</a></td>
</tr>
<tr>
<td>9</td>
<td>Jamie_Mason</td>
<td>#Baltimore Beer Week continues w/ a beer brkfst, beer pioneers luncheon, drink &amp; donate event, beer tastings &amp; more. <a href="http://ping.fm/VyTwg">http://ping.fm/VyTwg</a></td>
</tr>
<tr>
<td>10</td>
<td>carichardson</td>
<td>Seattle and Beer: I went to Seattle last weekend. It was my friend’s stag - he likes beer - we drank beer.. <a href="http://tinyurl.com/cpb4n9">http://tinyurl.com/cpb4n9</a></td>
</tr>
</tbody>
</table>
LiveTweet

Submit your tweet, and we tell you how interesting it is in the Twitter world right now:

your tweet ...

Tell me!

http://livetweet.west.uni-koblenz.de/
Compare Interestingness to Detect Trends

Tomorrow I am going to …

Interestiness:
- play tennis
- play golf
- do some gardening
- spread wisdom
- go to a Justin Bieber concert
- win the lottery
Community Simulation

Risk Mitigation: Simulating Effects of Policy Changes
• Governance of Communities
  – Def.: Steering and coordinating actions of community members

• Implementation:
  – Direct intervention of community owner
  – Functionality of the community platform

• Mitigation of risks:
  – Change platform functionality
  – Impact?
Governance by Policy Change

What if using Policy 2?

Policy 1

Community Manager

Prediction

Policy 2
Scenario: Policy Controlled Content Generation

- Users generate content
- Users search content
- Users consume content
- Users interact with content
  - Rate content
  - Reply to content

Where can a (platform) policy have effects?
Policy impacts on user behaviour

User Need for Content Consumption

Content Selection → Content Assessment → Rating → Replying

User Need for Content Creation

Content Creation → Content Evaluation

Policy
Attention Management

• Part of Community: forum activity
• Goal: Steer user activity to specific forums
• User model parameters:
  – Activity rate for creating threads
  – Activity rate for creating replies
  – Preferences for activity in specific forums

• Community Model
  – **Varied** restrictions for thread creation

• Observed Metrics
  – Response time on threads
Simulation based on observed parameters
Variable Parameters indicated by different policy

- Users search content
  - Presentation → Ranking threads in content views
    - Recency
    - Social Closeness
    - Topical closeness
    - Popularity
  - Observation: Influence which questions are answered

- Users generate content
  - Restrict number of questions asked per forum
    - Users turn to other questions
  - Observation: Response time in some fora reduced
Backend Technologies

Indexing Distributed Semantic Graphs
Community Information

• Examples
  – Male persons who have a public profile document
  – Computing science papers authored by social scientists
  – American actors who are also politicians and are married to a model.

• Maybe specific databases available:
  – Person search engines
  – Bibliographic databases
  – Movie database

How to integrate?
Linked Data

Semantic Web Technology to

1. Provide structured data on the web
2. Link data across data sources
SELECT ?x
WHERE {
  ?x rdfs:type foaf:Person .
  ?x rdfs:type pim:Male .
  ?x foaf:maker ?y .
  ?y rdfs:type
    foaf:PersonalProfileDocument .
}
SELECT ?x
WHERE {
  ?x rdfs:type foaf:Person .
  ?x rdfs:type pim:Male .
  ?x foaf:maker ?y .
  ?y rdfs:type
    foaf:PersonalProfileDocument .
}
SELECT ?x
FROM ...
WHERE {
  ?x rdfs:type foaf:Person .
  ?x rdfs:type pim:Male .
  ?x foaf:maker ?y .
  ?y rdfs:type
    foaf:PersonalProfileDocument .
}
How it works ...

pim:Male

foaf:Person

foaf:PPD

TC_18

5_76

TC_76

EQC_5

foaf:maker

hasEqC

hasEqC

hasDatasource

DS 1

DS 2

DS

DS

DS

RDF classes

type clusters

equivalence classes

data sources
Building the Index from a Stream

- Stream of data (coming from a LD crawler)

... D16, D15, D14, D13, D12, D11, D10, D9, D8, D7, D6, D5, D4, D3, D2, D1
Does it work good?

Comparison of stream based vs. Gold standard Schema on 11 M triple data set

![Bar chart comparing precision and recall for different cache sizes and methods: RDF-Class, TC, TC+S-TC, EQC, EQC+TC.](chart.png)

- **Cachesize:**
  - 100
  - 1k
  - 10k
  - 50k
  - 100k

**Winner BTC 2011**
Managing Business Communities

Lessons learned
Conclusions

• Business communities vary with regard to
  – Interaction
  – Interests
  – Type of conversation

• Novel analysis techniques needed:
  – Integration of different data sources
  – Simulation of policy changes
  – Value of users

• Concrete needs confirmed by project external companies looking for such technology
• S. Angeletou, M. Rowe, H. Alani. Modelling and analysis of user behaviour in online communities. In Proceedings of the 10th international conference on The semantic web - Volume Part I,


• M. Konrath, T. Gottron, A. Scherp. Schemex – web-scale indexed schema extraction of linked open data. In Semantic Web Challenge, Submission to the Billion Triple Track, 2011.


Thank You!