Introduction to my Ph. D. research, current and future directions

Research on Personalized Trust Evaluation in Online Social Networks

Presenter: Wenjun Jiang
Supervisor: Prof. Guojun Wang
Co-supervisor: Prof. Jie Wu

School of Information Science and Engineering, Hunan University    wenjj8a@163.com
Outline

Background

Generating Trusted Graphs for Trust Evaluation in OSNs

On Selecting Recommenders for Trust Evaluation in OSNs

FluidRating: Time-evolving Rating Prediction Scheme

Discussion
Background-OSN

A platform to build social networks or social relations among people who share interests, activities, backgrounds or real-life connections
Backgroud - Challenges in OSNs

Users:
- Construct connection
- Generate content

Relations:
- Topic related
- Time-sensitive

Information:
- Cross time and space
- Difficult to judge true or false

Trust Issues
- Interaction partner?
- Source of Information?
- Service quality?
Key Problem

When deciding if or not to interact with a stranger, “Can I trust the target?”
Key Components

- Source
- Target
- Evidence
- Influence

Providing a personalized and topic related fine-grained trust evaluation scheme
Main Contributions During Ph. D Studies

- **C1**: Trusted Graph Generation Based on Small-world Network Theory
- **C2**: Fine-grained feature-based social influence evaluation model
- **C3**: Trust evaluation using generalized maximum flow
- **C4**: Rating Scheme in Trust-based Recommendation Systems Using Fluid Dynamics

*trust is the basis*

*influence is the tool*

*human to item rating*

*human to human evaluation*
Trusted Graph Generation Based on Small-world Network Theory

1. Problem
2. Solution
3. Experiments and Analysis
4. Conclusion

SWTrust

SWTrust: Problem

How to generate trusted graphs and how to evaluate its quality
Solution

Key algorithm: trusted path searching based on user active domain

Basic theory: small world network, weak tie
Fig. 3. (a) The three categories of local neighbors, longer contact and longest contact; (b) Selection of next hop neighbors
Constructing trust network, generating trusted graph

• Search trustor-trustee paths
  – breadth first
  \[ p(i, j) = \frac{\lambda_1 \cdot x(j) + \lambda_2 \cdot y(j)}{\circ D_{topic} + \circ D_{target}} \]

• Define direct trust
  – recommendation trust
    \[ RT(i, j) = p(i, j) \]
  – functiona trust

• Further filtering
  – delete untrustful paths
  – delete too long paths
Experimental design

• Data set: www.epinions.com
  – web of trust
  – 51,888 edges, 3,168 nodes
• Evaluation method: Leave-one-out
• Two groups of experiments
  – Coverage
  – Accuracy
    • Absolute error, precision,
    • recall, Fscore
Experimental results - the coverage

Findings: when $k$ is smaller, the coverage decreases a little, but the efficiency improves significantly.
Experimental results - the accuracy

- Parameter setting: $L=\{2,8\}$, $brate=0.3$, $th=0.5$
Experimental results - the accuracy

• Parameter setting: L=6, brate=0.3, th=[0.1,0.9]
Summary - SWTrust

1. Extracting objective information of user active domain, selecting users' neighbors by category.

2. Emphasis of weak tie, considering both the topic and the target related degree.

3. Experimental validation of coverage and accuracy.
On Recommender Selection for Trust Evaluation in Online Social Networks

1. Problem
2. Solution
3. Experiments and Analysis
4. Conclusion

Introduction
- Online Social Networks

Facebook  Twitter  Google+  Tencent QQ  Sina WeiBo  Renren

Epinions 😊😊😊😊😊
Introduction
- Trust Issues & Trust Evaluation

**Trust issues** exist in any application whenever a person (e.g., source $s$) needs to estimate the trust level of another (e.g., target $t$), so as to decide whether or not to conduct further interactions.

**Trust Evaluation** is a process to predict the trust worthiness of a target $t$, from the perspective of $s$. 
Motivation
- From Social Graph to Trusted Graph

(a) Social Graph

\[ N_s = \{ u_1, \ldots, u_n \} \]

(b) Trusted Graph

\[ R = \{ r_1, \ldots, r_m \} \]
Motivation
- Select Whom As Recommenders?

The direct trust?
Social relationships?
Possible cost?
Risk (uncertainty)?
The Problem
- Recommender Selection Problem

**Given:** a social network $G = (V, E)$; Two nodes, $s$ and $t$, $s$ is the source and $t$ is the target.

**Find:** the best recommenders $R = \{r_1, ..., r_m\}$

**Objectives:** making a proper decision (to trust or not to trust $t$), meeting the optimal requirements of higher accuracy, lower risk (uncertainty), and less cost.
RATE Scheme
- Metrics Identification

**Trustworthiness** ($t_{uv}$): Honesty, and the capability to provide real information

**Influence** ($i_{uv}$): The closer the relationship exists between two persons, the larger the possibility that one’s opinion will influence the other’s.

**Uncertainty** ($u_{uv}$): It is lower, when the evidence for success/failure dominates, and it is larger when there is little or no evidence.

**Cost** ($c_{uv}$): Just as in daily life, the source wants to contact the target. Regardless of whether it contacts directly or indirectly, some cost will be charged.

$$M_{uv} = \langle t_{uv}, i_{uv}, u_{uv}, c_{uv} \rangle$$
RATE Scheme

- Utility Functions And The Objective

\[ F = w_t \times t + w_i \times i \quad (1) \]
\[ G = w_u \times u + w_c \times c \quad (2) \]

where \( w_t, w_i, w_u, w_c \) are the weights of the four metrics \( t, i, u, \) and \( c \), respectively (determined by the source \( s \)); \( 0 < w_t, w_i, w_u, w_c < 1, w_t + w_i = 1, w_u + w_c = 1. \)

The objective: maximize \( F \) and minimize \( G \)

Normalized Utility: \( \lambda * F + (1- \lambda) * (1-G) \) where \( 0 < \lambda < 1. \)
RATE Scheme

- 1-hop Recommender Selection

Issue 1: How to measure the quality of a recommender?

   measure the quality of a recommender

Issue 2: How many recommenders are enough, and are efficient for, decision-making?

   decide the size of the optimal recommender set
**RATE Scheme**

- Measure The Quality Of A Recommender

*Quality of Recommender (QoR)* comprises requirements on a recommender, taking trustworthiness, influence, uncertainty, and cost, as attributes.

Users can set multiple quality constraints \( Q^t, Q^i, Q^u, Q^c \) (e.g., thresholds)
RATE Scheme

- Example

\[ Q_t > 0.5, \, Q_i > 0.5, \, Q_u < 0.3, \, \text{Qualified recommenders: } u_1, \, u_2 \]
RAT£ Scheme
- The Size Of The Recommender Set

Selecting all qualified neighbors.

Selecting a fixed number of qualified neighbors
e.g., 3, 6, etc.

Selecting a fixed proportion of qualified neighbors
e.g., 1/3, 1/6, etc.
RAGE Scheme
- The Size Of The Recommender Set

Flexibly selecting some top $m$ qualified neighbors, $m \leq n$.

We continue to select qualified recommenders until the number of next hop neighbors is no less than the current ones.
RATE Scheme

- The Effects Of RATE

By comparing the performance of sorting or not sorting the neighbors with QoR, the effects of RATE scheme can be analyzed.
Extension
- Multiple Hop Scenario

Trustworthiness

\[ t_{p(a_1,\ldots,a_n)} = \prod_{e(a_j,a_{j+1})\in p(a_1,\ldots,a_n)} t_{a_j,a_{j+1}} \]

Influence

\[ i_{p(a_1,\ldots,a_n)} = \prod_{e(a_j,a_{j+1})\in p(a_1,\ldots,a_n)} i_{a_j,a_{j+1}} \]

Uncertainty

\[ u_{p(a_1,\ldots,a_n)} = 1 - \prod_{e(a_j,a_{j+1})\in p(a_1,\ldots,a_n)} (1 - u_{a_j,a_{j+1}}) \]

Cost

\[ c_{p(a_1,\ldots,a_n)} = \sum_{e(a_j,a_{j+1})\in p(a_1,\ldots,a_n)} c_{a_j,a_{j+1}} \]
Experimental Evaluation
- Dataset Epinions (www.epinions.com)
Experimental Evaluation
- Method: Leave-One-Out

If there is an edge between two nodes, that edge is masked, and trust is calculated through algorithms; then, we compare the calculated value with the masked value.
Experimental Evaluation
- Metrics: Precision, Recall, Fscore

Precision = \frac{A_t \cap B_t}{B_t}

Recall = \frac{A_t \cap B_t}{A_t}

Fscore = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}

where \(A_t\) is the number of edges on which \(s\) trusts \(t\) directly, and \(B_t\) is the number of edges on which \(s\) trusts \(t\), by trust calculated through an algorithm.
Experimental Evaluation
- Trust Evaluation Strategies

AveR-MaxT
AveR-WAveT
MaxR-MaxT
MaxR-WAveT
Experimental Evaluation
- Results (Accuracy)
Evaluation
- Results (Cost & Uncertainty)
Summary - RATE

1. Identify the problem of recommender selection
2. Propose a recommendation-aware trust evaluation scheme: RATE
3. Evaluate RATE using a real trust network, Epinions.

Wenjun Jiang\textsuperscript{a,b}, Prof. Jie Wu\textsuperscript{b}, Prof. Guojun Wang\textsuperscript{a} and Huanyang Zheng\textsuperscript{b}

\textsuperscript{a} Central South University
\textsuperscript{b} Temple University
Outline

1. Trust-based Recommendation
2. Motivation
3. Problem
4. Solution: FluidRating
5. Experimental Evaluation
6. Conclusion & Future Work
Personalized Recommendation

Too many choices in daily life: which restaurant for dinner, which movie to watch, which product to purchase....

Restaurant near Philadelphia, PA

- **Zahav**: 4.5 stars (565 reviews) - Israeli
- **Five Guys Burgers and Fries**: 3.9 stars (267 reviews) - Hamburger
- **Raw Sushi & Sake Lounge**: 4.2 stars (213 reviews) - Sushi
- **Bibou**: 4.6 stars (239 reviews) - French
- **El Vez**: 4.2 stars (1,118 reviews) - Mexican
- **Ruth's Chris Steak House**: 4.3 stars (484 reviews) - $$$$ - Steak
- **Butcher and Singer**: 4.5 stars (379 reviews) - $$$ - Steak
1. Trust-based Recommendation

- Two types of information
  - user to user: web of trust
  - user to item: review, rating

- Basic idea
  - use the knowledge of a trust network among users, to provide personalized recommendations by aggregating the opinions of their trusted friends
2. Motivation

- Consider time evolving effects
  - users receive different influences at different times
  - upon receiving influence, users can react variously
- Differentiate direct and indirect influences
- Capture user features
  - key feature we are considering: persistency
  - how much one insists on his opinion

High quality personalized recommendation
2. Motivation

- Existing trust-based recommendation methods
  - calculate at the current time
  - take direct friends and friends of friends equally
  - assume adoption of all influence

- In real life
  - time-evolving system
  - closer friends have more chance of influence
  - upon receiving influence, different users may take different actions (depends on user feature)
Problem
System setting: Rating network

- **Nodes**
  - raters \( R = \{a_1, a_2\} \)
  - non-raters \( N = \{a_3, a_4, a_5\} \)
  - sink \( S = \{a_6\} \)

- **Influence relations**
  - converted from trust relations
  - from raters to sink
  - from raters to non-raters, then to sink
Problem
Time-evolving rating prediction

Tasks
- predict rating efficiently
- reflect time-evolving effect
- capture user feature
Solution

- Time-evolving opinion formulation process
  - each user receives the influence
  - updates his own opinion
  - propagates his opinion to other friends

- Discretized view
  - each user exchanges opinions with his neighbors

- Model the process using fluid dynamics theory
Solution

- **Social principles**
  - Principle 1: First Influence Dominates.
  - Principle 2: Stronger Influence Dominates.

- **Physical principles**
Solution

- **FluidRating**: three components
  - **container**: user
  - **pipe**: influence relation
  - **fluid**: recommendation
    - temperature as rating
    - height as persistency

- **Influence**: two micro steps
  - compare persistency (fluid height)
  - fluid flowing from one container to another
Solution

FluidRating
- from rating network to fluid dynamics system

rating network  fluid dynamics system
Solution

- **FluidRating**: 3 steps
  - Fluid Updating Preparation
    - Calculate the fluid volume that will flow *(each pair of users)*
  - Fluid Updating Execution
    - Let fluid flow and mix *(all users)*
  - Sample Aggregation
    - Collect and aggregate samples *(from multiple rounds)*
Solution

Step 1: Fluid Updating Preparation

- The speed of efflux: Torricelli's law
  \[ v_{aa'} = \sqrt{2g(h_a - h_{a'})} \]

- The volume of fluid that will flow
  \[ s_{aa'} = \sqrt{2g[h_a(i) - h_{a'}(i)]} \cdot w_{aa'} \cdot \Delta \]

- The temperature of fluid that will flow
  \[ t_{aa'} = t_a \]
Solution

Step 2: Fluid Updating Execution

- the updated volume

\[
s_a(i + 1) = s_a(i) - \sum_{a' \in N_a^{out}} s_{aa'} + \sum_{a'' \in N_a^{in}} S_{a''a}
\]

- the updated temperature

\[
t_a(i + 1) = \frac{t_a(i) \cdot [s_a(i) - \sum_{a' \in N_a^{out}} s_{aa'}] + \sum_{a'' \in N_a^{in}} t_{a''a} \cdot S_{a''a}}{s_a(i + 1)}
\]
Step 2: Fluid Updating Execution

Example

<table>
<thead>
<tr>
<th>Time slot</th>
<th>Containers</th>
</tr>
</thead>
<tbody>
<tr>
<td>k=0</td>
<td>a₁ a₂ a₃ a₄ a₅ a₆ a₀</td>
</tr>
<tr>
<td>k=1</td>
<td>a₁ a₂ a₃ a₄ a₅ a₆ a₀</td>
</tr>
<tr>
<td>...</td>
<td>a₁ a₂ a₃ a₄ a₅ a₆ a₀</td>
</tr>
<tr>
<td>k=16</td>
<td>a₁ a₂ a₃ a₄ a₅ a₆ a₀</td>
</tr>
<tr>
<td>k=17</td>
<td>a₁ a₂ a₃ a₄ a₅ a₆ a₀</td>
</tr>
</tbody>
</table>

Pipes
Solution

Step 3: sample aggregation

- aggregation sequence can be uniform or non-uniform
- give earlier samples more weight

\[
t_{a_n} = \sum_{i=1}^{k} q^{1+c(i-1)} \cdot t_{a_n}(i)
\]
The height and temperature become stable after some time.
The variation decreases with iterations.
Experimental Evaluation

- **Data set:** Epinions
- **Test method:** Leave-one-out
- **Metric:** RMSE  \[
RMSE = \sqrt{\frac{\sum (r_{u,i} - \hat{r}_{u,i})^2}{D}}
\]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Default value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$</td>
<td>fluid height in rater’s container</td>
<td>10</td>
</tr>
<tr>
<td>$b$</td>
<td>cross-sectional area of containers</td>
<td>1</td>
</tr>
<tr>
<td>$k$</td>
<td>number of iterations</td>
<td>250</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>time slot</td>
<td>0.04</td>
</tr>
<tr>
<td>$c$</td>
<td>(non-)uniform aggregation</td>
<td>(1)0</td>
</tr>
<tr>
<td>$q$</td>
<td>uniform aggregation</td>
<td>$1/k$</td>
</tr>
<tr>
<td></td>
<td>nonuniform aggregation</td>
<td>[0.1,0.9]</td>
</tr>
</tbody>
</table>
Experimental Evaluation

- The effects of first influence

(a) and (b) show four different patterns of 4 user/item test pairs. In all the four patterns, the first samples give predictions close to the real truth.

(c) provides a comparison of the average ratings with respect to the number of iteration (i.e., $k$) and the maximum length.
Experimental Evaluation

- The effects of impact factors

The left figure depicts the average predicted rating with FluidRating 1 and different settings of the time slot duration. The right figure shows the RMSE of FluidRating 1 with $c=1$ and $q$ changing from 0.1 to 0.9.
Experimental Evaluation

The effects of aggregation methods

This figure compares FluidRating 1 with $c = 0$, $c = 1$, and $q = 0.5$, the latter is denoted as FluidRating 1$^*$. 

Finding: When we put more weight on the earlier influence, the accuracy is improved.
Experimental Evaluation

Comparison of Multiple Methods

The comparison of several trust-based recommendation methods

Finding: FluidRating beats others; the RMSE of using FluidRating 1* is 4.812% less than that of using TidalTrust when maximum length=6
Summery of Experiments

- Validate the time-evolving effects
- Validate the existence of first influence
- Test the effects of several factors
  - iteration number
  - time duration
  - aggregation sequence
  - sample approach (see details in paper)
Future Work for RATE

The theoretical bounds of the size of an optimal recommender subset.

The probability of success to make a proper trust decision.
Future Work for FLuidRating

Conclusion

- FluidRating can reflect the time-evolving feature
- Differentiate direct and indirect influence
- Reflect the user personality feature (persistency)

Future work

- More personality feature (e.g., persuasiveness)
- Real experience evolution
Discussion: How to select a research problem

- Simple definition
- Elegant solution
- Room for imagination

http://www.cis.temple.edu/~wu/

Discussion: How to write a research paper

• 10-100-1000-10000
  – Editor's Note: How to Write Research Articles in Computing and Engineering Disciplines

• More methods
  – http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWtoc.html
  – http://www.aresearchguide.com/1steps.html
Some suggestions

• Paper reading
  – Top conference/journal
  – http://www.ccf.org.cn/sites/ccf/paiming.jsp

• Paper editing
  – latex
  – http://www.latex-project.org/

• Figure drawing
  – microsoft visio

• Experimental resulting figure drawing
  – gnuplot
  – http://www.gnuplot.info/
Thank you for your attentions!

Wenjun Jiang
Assistant professor in HU
Email: wenjj8a@163.com