Measuring User Influence in Financial Microblogs: Experiments Using StockTwits Data

Paulo Cortez
pcortez@dsi.uminho.pt

Nuno Oliveira
nunomroliveira@gmail.com

João P. Ferreira
a61572@alunos.uminho.pt

WIMS’16 - 6th Int. Conference on Web Intelligence, Mining and Semantics.
Measuring User Influence in Financial Microblogs, Cortez et al., WIMS'16
Motivation (I)

• There is a growing interest in the identification of influential users in social networks (e.g. Klout).
• Such identification can allow a better understanding of dominant influential trends, which can be used in decision support (e.g. more effective marketing).
• Using social media analytics (e.g. sentiment analysis) to model and forecast stock market behavior is a promising research topic.
• Financial microblog data (e.g. Twitter, StockTwits) is easy to collect and the financial community that uses such platforms has grown.
Motivation (II)

• Research that combines user influence measures and financial social media analytics is scarce.
• Most financial social media forecasting methods simply aggregate all messages from all users.
• Filtering or weighting such messages according to a user influence criterion might lead to better sentiment indexes and forecasts.
• The automatic identification of influential users is also relevant for financial media platforms (e.g. suggestion of most “interesting” users).
In this paper

- We explore three known social network measures (*indegree*, *betweenness*, *page rank*) and also test a new *posts* measure, to identify relevant users.
- We use a very large dataset (with 1.2 million messages) from StockTwits, a popular financial microblog.
- We consider direct interactions between two users to build social network graphs.
- User rankings are evaluated using a robust rolling windows with five consecutive evaluation periods.
- Two criteria are used:
  - Capacity to select quality users;
  - Daily sentiment correlation with other users.
StockTwits Data: labeled messages

- Microblogging platform exclusively dedicated to **stock market**
  - Less noisy when compared with generalist microblogging services
  - Sample messages: $VIX$ turns negative on the day
    $EVRY$ - http://stks.co/g0mqq - Tidal Wave Volume - Currently at 5 x average daily volume.

- Sentiment metadata:
  - Users are able to classify their messages as **bullish** or **bearish**
  - Data from June 2010 until March 2013 (341,230 labeled messages; much higher than classified data sets applied in most studies about mining finance microblogging data).

- Messages can be filtered by their cash tag. In this paper, we selected:
  - **ALL** (no selection);
  - **$AAPL** (Apple);
  - **$GOOG** (Google).
StockTwits Data: labeled messages

Figure 1: Histogram of StockTwits messages over the time periods analyzed and for the ALL, AAPL and GOOG selections (x-axis denotes the time period $t_i$; y-axis the frequency of the messages)
StockTwits Data: quality users

• We had access to a static dataset with user information at March 2013.

• The static dataset contains a set of 300 users ($S_U$) that were labeled by StockTwits as “suggested” (high quality contributors, included in a curated list provided for normal users).
• During the period of analysis, StockTwits users could perform several operations, namely: retweet, share (special button) and reply (generating a conversation).
• We assume there is a direct interaction between user A and B (A → B) if there is a retweet, share or reply from user B to a message posted by user A.
• We used all these direct interactions (except self interactions) to generate direct graphs $G_{t_i}$ (build using data from time period $t_i$).
Figure 2: Histogram of StockTwits messages over the time periods analyzed and for retweets, shares and replies ($x$-axis denotes the time period $t_i$; $y$-axis the frequency of the messages)

Measuring User Influence in Financial Microblogs, Cortez et al., WIMS'16
StockTwits Data: social network graphs

Figure 3: StockTwits social network graph (with minimum of 10 interactions) $G_0$ (left) and $G_4$ (right)

Measuring User Influence in Financial Microblogs, Cortez et al., WIMS'16
Four User Influence Measures

• **Indegree** – number of users that interact with user A.
• **Betweenness** – indicator of network centrality.
• **Page Rank** – based on the famous Google search.
• **Posts** - number of posts from user A that received a direct interaction (retweet, share or reply).
List of Top Users

• All users were ranked according to a user influence measure.

• For each user measure, we define five lists of top users: Top 12 (T12), Top 25 (T25), Top 50 (T50), Top 100 (T100) and Top 200 (T200).

<table>
<thead>
<tr>
<th>User</th>
<th>Indegree</th>
<th>Betweenness</th>
<th>Page Rank</th>
<th>Posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>2</td>
<td>167.5</td>
<td>0.011</td>
<td>67</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>0.0</td>
<td>0.012</td>
<td>20</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>175.0</td>
<td>0.013</td>
<td>13</td>
</tr>
<tr>
<td>d</td>
<td>3</td>
<td>61.0</td>
<td>0.017</td>
<td>41</td>
</tr>
<tr>
<td>e</td>
<td>5</td>
<td>171.7</td>
<td>0.020</td>
<td>109</td>
</tr>
</tbody>
</table>

The top list T3 for Indegree includes users \{e, b, d\}
Evaluation

• **Rolling window** – data split into 6 periods (each with 6 months except last one):
  
  • Training data used to build the social network graphs.
  
  • Test data used to evaluate the user measures.

<table>
<thead>
<tr>
<th>Time</th>
<th>$N_G$</th>
<th>Messages</th>
<th>Begin</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_0$</td>
<td>1336</td>
<td>14729</td>
<td>June 2, 2010</td>
<td>December 1, 2010</td>
</tr>
<tr>
<td>$t_1$</td>
<td>1755</td>
<td>31924</td>
<td>December 2, 2010</td>
<td>June 1, 2011</td>
</tr>
<tr>
<td>$t_2$</td>
<td>2880</td>
<td>31983</td>
<td>June 2, 2011</td>
<td>December 1, 2011</td>
</tr>
<tr>
<td>$t_3$</td>
<td>3851</td>
<td>30667</td>
<td>December 2, 2011</td>
<td>June 1, 2012</td>
</tr>
<tr>
<td>$t_4$</td>
<td>12414</td>
<td>80865</td>
<td>June 2, 2012</td>
<td>December 1, 2012</td>
</tr>
<tr>
<td>$t_5$</td>
<td>–</td>
<td>150463</td>
<td>December 2, 2012</td>
<td>March 30, 2013</td>
</tr>
</tbody>
</table>
Evaluation

• Evaluation Criteria:
  • Percentage of Quality Users (PQU).
  • Sentiment Analysis Correlation: Spearman’s rank correlation values between the bullishness indexes for the Top List users and other users.
Experiments

• All operations were executed using the open source R tool (www.r-project.org).

• We analyze a total of $4 \times 5 = 20$ lists of top users.
## Results: Quality Users

Table 3: Capacity to select quality users (PQU values, in %, best value per period is underlined, values $\geq 50\%$ are in bold)

<table>
<thead>
<tr>
<th>Time</th>
<th>Indegree T12 T25 T50 T100 T200</th>
<th>Betweenness T12 T25 T50 T100 T200</th>
<th>Page Rank T12 T25 T50 T100 T200</th>
<th>Posts T12 T25 T50 T100 T200</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>75 84 76 64 52</td>
<td>83 72 44 46 30</td>
<td>75 84 76 63 50</td>
<td>67 76 74 66 53</td>
</tr>
<tr>
<td>$t_2$</td>
<td>67 80 78 72 56</td>
<td>83 76 54 46 39</td>
<td>75 80 80 69 55</td>
<td>75 80 80 72 55</td>
</tr>
<tr>
<td>$t_3$</td>
<td>58 76 78 69 54</td>
<td>58 56 56 43 34</td>
<td>67 80 78 67 54</td>
<td>58 76 80 71 56</td>
</tr>
<tr>
<td>$t_4$</td>
<td>25 16 16 16 15</td>
<td>33 36 26 27 18</td>
<td>17 20 16 14 18</td>
<td>33 24 18 21 22</td>
</tr>
<tr>
<td>$t_5$</td>
<td>25 20 18 12 10</td>
<td>25 28 20 14 12</td>
<td>25 20 22 13 12</td>
<td>25 20 20 13 12</td>
</tr>
</tbody>
</table>

Avg. 50 55 53 47 38 57 54 40 35 27

O.A. 48 42 49 50

Measuring User Influence in Financial Microblogs, Cortez et al., WIMS'16
# Results: Sentiment Analysis Correlation

Table 4: Correlation sentiment analysis values (best value per period is underlined; values ≥ 0.40 are in bold)

<table>
<thead>
<tr>
<th>Time</th>
<th>( N_M )</th>
<th>( N )</th>
<th>( T_{12} )</th>
<th>( T_{25} )</th>
<th>( T_{50} )</th>
<th>( T_{100} )</th>
<th>( T_{200} )</th>
<th>( T_{12} )</th>
<th>( T_{25} )</th>
<th>( T_{50} )</th>
<th>( T_{100} )</th>
<th>( T_{200} )</th>
<th>( T_{12} )</th>
<th>( T_{25} )</th>
<th>( T_{50} )</th>
<th>( T_{100} )</th>
<th>( T_{200} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>861 177</td>
<td>0.42</td>
<td>0.34 N</td>
<td>0.38 N</td>
<td>0.47 N</td>
<td>0.48 N</td>
<td>0.27 N</td>
<td>0.33 N</td>
<td>0.35 N</td>
<td>0.43 N</td>
<td>0.45 N</td>
<td>0.39 N</td>
<td>0.40 N</td>
<td>0.38 N</td>
<td>0.44 N</td>
<td>0.49 N</td>
<td>0.43 N</td>
</tr>
<tr>
<td>( t_1 )</td>
<td>1207 184</td>
<td>0.33 N</td>
<td>0.62 N</td>
<td>0.58 N</td>
<td>0.60 N</td>
<td>0.54 N</td>
<td>0.54 N</td>
<td>0.58 N</td>
<td>0.60 N</td>
<td>0.57 N</td>
<td>0.61 N</td>
<td>0.61 N</td>
<td>0.63 N</td>
<td>0.64 N</td>
<td>0.64 N</td>
<td>0.59 N</td>
<td>0.57 N</td>
</tr>
<tr>
<td>( t_3 )</td>
<td>1407 184</td>
<td>0.58 N</td>
<td>0.61 N</td>
<td>0.58 N</td>
<td>0.60 N</td>
<td>0.47 N</td>
<td>0.49 N</td>
<td>0.59 N</td>
<td>0.59 N</td>
<td>0.58 N</td>
<td>0.63 N</td>
<td>0.60 N</td>
<td>0.63 N</td>
<td>0.62 N</td>
<td>0.63 N</td>
<td>0.60 N</td>
<td>0.58 N</td>
</tr>
<tr>
<td>( t_4 )</td>
<td>4560 171</td>
<td>0.36 N</td>
<td>0.40 N</td>
<td>0.45 N</td>
<td>0.46 N</td>
<td>0.40 N</td>
<td>0.40 N</td>
<td>0.45 N</td>
<td>0.46 N</td>
<td>0.58 N</td>
<td>0.41 N</td>
<td>0.39 N</td>
<td>0.43 N</td>
<td>0.41 N</td>
<td>0.45 N</td>
<td>0.43 N</td>
<td>0.40 N</td>
</tr>
<tr>
<td>( t_5 )</td>
<td>6057 119</td>
<td>0.08 N</td>
<td>0.14 N</td>
<td>0.14 N</td>
<td>0.07 N</td>
<td>-0.02 N</td>
<td>0.03 N</td>
<td>0.02 N</td>
<td>0.21 N</td>
<td>0.30 N</td>
<td>0.02 N</td>
<td>0.11 N</td>
<td>0.22 N</td>
<td>0.12 N</td>
<td>0.23 N</td>
<td>0.06 N</td>
<td>0.13 N</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.35 N</td>
<td>0.42 N</td>
<td>0.41 N</td>
<td>0.43 N</td>
<td>0.43 N</td>
<td>0.34 N</td>
<td>0.37 N</td>
<td>0.39 N</td>
<td>0.46 N</td>
<td>0.50 N</td>
<td>0.41 N</td>
<td>0.42 N</td>
<td>0.46 N</td>
<td>0.45 N</td>
<td>0.47 N</td>
<td>0.41 N</td>
<td>0.42 N</td>
</tr>
<tr>
<td>O.A.</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
</tr>
</tbody>
</table>
Results: Sentiment Analysis Correlation

ALL t2 Page Rank T50 (cor.=0.64, 184 points)

ALL t4 Betweenness T200 (cor.=0.58, 171 points)

Measuring User Influence in Financial Microblogs, Cortez et al., WIMS'16
Conclusions

• The best **Percentage of Quality Users** (PQU) was obtained by the **posts** measure (50%), followed by **page rank** (49%).

• PQU results for the last time periods suggest that the StockTwits **curated list** of quality users is **too conservative** and should be updated (300 vs 300,000 users!).

• **Page rank** achieved the best overall sentiment analysis correlation results (0.44 for ALL, 0.38 for AAPL, 0.28 for GOOG), followed by the **posts** measure.

• Some specific top lists obtained interesting overall results: **betweenness T200** for ALL (0.50), **page rank T25** and **posts T50** for AAPL (0.41).
Ongoing Work

• **PhD Thesis** of Bruno Vieira (in preparation), where we expect to explore (among other possibilities):
  • Use of **other features** (e.g. number of likes);
  • Study **more complex influence effects** (e.g. diffusion of sentiment message cascades, the effect of time).
  • Use of similar user measures to **weight the sentiment opinions** when designing forecasting models of stock market variables (e.g., returns, volatility).
Acknowledgments

• We wish to thank StockTwits (www.stocktwits.com) for kindly providing their data.

• Work supported by COMPETE: POCI-01- 0145-FEDER-007043 and FCT within the Project Scope: UID/CEC/00319/2013.